ORGANIC CHEMISTRY

Good reasons must, of force, give place to better.

JULIUS CÆSAR

Titles in the E.L.B.S. Low-priced Textbook Series

Emeleus and Anderson, *Modern Aspects of Inorganic Chemistry* 12s
Finar, *Organic Chemistry* Volumes I and II, 12s each
Kipping and Kipping, *Organic Chemistry* Parts I and II, 9s Part III, 9s
Mee, Physical Chemistry. 12s
Partington, *A Textbook of Inorganic Chemistry* 12s
Vogel, *A Textbook of Quantitative Inorganic Analysis* 21s
Besant and Ramsey, *A Treatise on Hydromechanics.* Parts I and II, 12s
Ramsey, *Dynamics* Parts I and II, 9s each
Starling, *Electricity and Magnetism for Degree Students.* 12s.
Blakey, *University Mathematics* 12s
Copson, *An Introduction to the Theory of Functions of a Complex Variable* 12s
Hardy, *A Course of Pure Mathematics* 9s
Littlewood, *A University Algebra* 12s
Titchmarsh, *The Theory of Functions* 12s
Weatherburn, *A First Course in Mathematical Statistics.* 12s
Weatherburn, *Elementary Vector Analysis* 9s.
Baker, *The Steel Skeleton* Volumes I and II 15s each
Bevan, *The Theory of Machines* 12s
Blyth, *A Geology for Engineers* 12s
Cattermole, *Transistor Circuits* 15s
Chapman, *Workshop Technology* Parts I and II, 9s each Part III, 12s
Cotton, *Electrical Technology.* 9s.
Cotton, *The Transmission and Distribution of Electrical Energy* 12s.
Gavin and Houldin, *Principles of Electronics* 12s
Golding, *Electrical Measurements and Measuring Instruments* 15s.
Kearton, *Steam Turbine Theory and Practice* 15s
Lewitt, *Hydraulics and Fluid Mechanics* 12s
Morley, *Theory of Structures* 12s.
Morley, *Strength of Materials* 12s
Rose, *Mathematics for Engineers* Parts I and II, 9s each
Say, *The Performance and Design of A C. Machines* 12s
Wrangham, *The Theory and Practice of Heat Engines* 15s
Crowther, *An Outline of Money* 9s
Marshall, *Principles of Economics.* 12s
Pigou, *The Economics of Welfare* 12s.
Stonier and Hague, *A Textbook of Economic Theory* 12s

ORGANIC CHEMISTRY

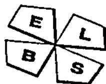
VOLUME TWO
STEREOCHEMISTRY AND THE CHEMISTRY OF NATURAL PRODUCTS

by

I. L. FINAR

B.Sc., Ph.D.(Lond.), A.R.I.C.

Senior Lecturer in Organic Chemistry,
Northern Polytechnic, Holloway, London



THE ENGLISH LANGUAGE BOOK SOCIETY

and

LONGMANS, GREEN & CO LTD

LONGMANS, GREEN AND CO LTD
48 GROSVENOR STREET, LONDON W1
RAILWAY CRESCENT, CROYDON, VICTORIA, AUSTRALIA
AUCKLAND, KINGSTON, (JAMAICA), LAHORE, NAIROBI
LONGMANS SOUTHERN AFRICA (PTY) LTD
THIBAULT HOUSE, THIBAULT SQUARE, CAPE TOWN
JOHANNESBURG, SALISBURY
LONGMANS OF NIGERIA LTD
W.R. INDUSTRIAL ESTATE, IKEJA
LONGMANS OF GHANA LTD
INDUSTRIAL ESTATE, RING ROAD SOUTH, ACCRA
LONGMANS GREEN (FAR EAST) LTD
443 LOCKHART ROAD, HONG KONG
LONGMANS OF MALAYA LTD
44 JALAN AMPANG, KUALA LUMPUR
ORIENT LONGMANS LTD
CALCUTTA, BOMBAY, MADRAS
DELHI, HYDERABAD, DACCA
LONGMANS CANADA LTD
137 BOND STREET, TORONTO 2

First Published 1956
New Impression 1958
Second Edition 1959
New Impression 1960
This edition by photolithography 1962

Printed in Great Britain by Butler & Tanner Ltd, Frome and London

PREFACE TO SECOND EDITION

THIS volume has now been revised to bring it up to date; this has involved the expansion of some sections and the addition of new material. It may be useful if I indicate briefly the more important changes I have made in this new edition. Two major additions are conformational analysis and biosynthesis: in each case I have given an introduction to the problem, and have also discussed various applications. Some other additions are nuclear magnetic resonance, correlation of configurations, *iso*flavones, and vitamin B_{12}. Expanded topics include dipole moments, molecular rotation, optical isomerism, steric effects (including steric factors and the transition state, molecular overcrowding), ascorbic acid, structure and synthesis of cholesterol, vitamin A_1, polypeptides, mechanism of enzyme action, flavones, streptomycin and patulin.

I wish to thank those reviewers and correspondents who have pointed out errors and have made suggestions for improving the book.

1958 I. L. FINAR

PREFACE TO FIRST EDITION

IN the Preface of my earlier book, *Organic Chemistry*, Longmans, Green (1954, 2nd ed.), I expressed the opinion that the chemistry of natural products is the application of the principles of Organic Chemistry. The present work is, in this sense, a continuation of my earlier one. It is my belief that a student who has mastered the principles will be well on the road to mastering the applications when he begins to study them. At the same time, a study of the applications will bring home to the student the dictum of Faraday: "Ce n'est pas assez de savoir les principes, il faut savoir *Manipuler*" (quoted by Faraday from the *Dictionnaire de Trevoux*).

In the sections on Stereochemistry, I have assumed no previous knowledge of this subject. This has meant a certain amount of repetition of some of the material in my earlier book, but I thought that this way of dealing with the subject would be preferable, since the alternative would have led to discontinuity. I have omitted

an account of the stereochemistry of co-ordinated compounds since this subject is dealt with in text-books on Inorganic Chemistry.

The section of this book dealing with natural products has presented many difficulties. I have tried to give a general indication of the problems involved, and in doing so I have chosen, to a large extent, the most typical compounds for fairly detailed discussion. At the same time, I believe that the subject matter covered should serve as a good introduction to the organic chemistry required by students reading for Part II of the Special Honours degree in chemistry of the London University. I have given a selected number of reading references at the end of each chapter to enable students to extend their knowledge and also to make up for any omissions I may have made. It is impossible to express my indebtedness to those authors of monographs, articles, etc., from which I have gained so much information, and I can only hope that some measure of my gratitude is expressed by the references I have given to their works.

Since physical measurements are now very much used in elucidating structures of organic compounds, I have included a short chapter on these measurements (Chapter I). I have introduced only a minimum amount of theory in this chapter to enable the student to understand the terms used; the main object is to indicate the *applications* of physical measurements.

In this book, cross-references are indicated by section and chapter. If a cross-reference occurs to another section in that chapter, then only the section number is given. It should also be noted that the numbers assigned to formulae, etc., are confined to each section, and not carried on to subsequent sections in that chapter. When references have been given to my earlier volume, the latter has been referred to as Volume I. In such cases the pages have not been quoted since the pagination of the various editions changes. The student, however, should have no difficulty in locating the reference from the index of Volume I.

I. L. FINAR

1955

LIST OF JOURNAL ABBREVIATIONS

ABBREVIATIONS	JOURNALS
Ann. Reports (Chem. Soc.)	Annual Reports of the Progress of Chemistry (The Chemical Society, London).
Ber.	Berichte der deutschen chemischen Gesellschaft (name now changed to Chemische Berichte).
Bull. Soc. chim.	Bulletin de la Société chimique de France.
Chem Reviews	Chemical Reviews.
Chem. and Ind.	Chemistry and Industry.
Experientia	Experientia.
Ind. chim. belg.	Industrie chimique belge.
Ind. Eng. Chem.	Industrial and Engineering Chemistry.
J. Amer. Chem. Soc.	Journal of the American Chemical Society
J. Chem. Educ.	Journal of Chemical Education.
J.C.S.	Journal of the Chemical Society.
J. Pharm. Pharmacol.	Journal of Pharmacy and Pharmacology.
Nature	Nature.
Proc. Leeds Phil. Lit. Soc., Scientific Section	Proceedings of the Leeds Philosophical and Literary Society, Scientific Section.
uart. Reviews (Chem. Soc.)	Quarterly Reviews of the Chemical Society (London).
ence	Science.

CONTENTS

CHAPTER I

PHYSICAL PROPERTIES AND CHEMICAL CONSTITUTION

§1. Introduction. A tremendous amount of work has been and is being done to elucidate the relationships between physical properties and chemical structure. An ideal state to be achieved is one where the chemist can predict with great accuracy the physical properties of an organic compound whose structure is known, or formulate the correct structure of an organic compound from a detailed knowledge of its physical properties. A great deal of progress has been made in this direction as is readily perceived by examining the methods of elucidating structures of organic compounds over the last few decades. In the early work, the structure of an organic compound was solved by purely chemical means. These are, briefly:

(i) Qualitative analysis.

(ii) Quantitative analysis, which leads to the empirical formula.

(iii) Determination of the molecular weight, which leads to the molecular formula.

(iv) If the molecule is relatively simple, the various possible structures are written down (based on the valency of carbon being four, that of hydrogen one, oxygen two, etc.). Then the reactions of the compound are studied, and the structure which best fits the facts is chosen. In those cases where the molecules are not relatively simple, the compounds are examined by specific tests to ascertain the nature of the various groups present (see, *e.g.*, alkaloids, §4. XIV). The compounds are also degraded and the smaller fragments examined. By this means it is possible to suggest a tentative structure.

(v) The final stage for elucidation of structure is synthesis, and in general, the larger the number of syntheses of a compound by *different* routes, the more reliable will be the structure assigned to that compound.

In recent years, chemists are making increasing use of physical properties, in addition to purely chemical methods, to ascertain the structures of new compounds. Furthermore, information on structure has been obtained from physical measurements where such information could not have been obtained by chemical methods. The early chemists identified pure compounds by

physical characteristics such as boiling point, melting point, refractive index; nowadays many other physical properties are also used to characterise pure compounds.

The following account describes a number of relationships between physical properties and chemical constitution, and their application to the problem of elucidating chemical structure.

§2. Van der Waals forces. Ostwald (1910) classified physical properties as **additive** (these properties depend only on the nature and number of atoms in a molecule), **constitutive** (these properties depend on the nature, number and arrangement of the atoms in the molecule), and **colligative** (these properties depend only on the number of molecules present, and are independent of their chemical constitution). It is extremely doubtful whether any one of these three classes of properties is absolutely independent of either or both of the others, except for the case of molecular weights, which may be regarded as truly additive and independent of the other two. In constitutive and colligative properties, forces between molecules have a very great effect on these properties. Attractive forces between molecules of a substance must be assumed in order to explain cohesion in liquids and solids. Ideal gases obey the equation $PV = RT$, but real gases do not, partly because of the attractive forces between molecules. Van der Waals (1873) was the first to attempt to modify the ideal gas law for the behaviour of real gases by allowing for these attractive forces (he introduced the term a/v^2 to correct for them). These intermolecular forces are now usually referred to as *van der Waals forces*, but they are also known as *residual* or *secondary valencies*. These forces may be forces of attraction or forces of repulsion; the former explain cohesion, and the latter must be assumed to exist at short distances, otherwise molecules would collapse into one another when intermolecular distances become very small. The distances to which atoms held together by van der Waals forces can approach each other, *i.e.*, the distances at which the repulsion becomes very large, are known as *van der Waals radii*. Some values (in Ångstroms) are:

H, 1·20; O, 1·40; N, 1·50; Cl, 1·80; S, 1·85.

These values are very useful in connection with molecules that exhibit the spatial effect, *e.g.*, substituted diphenyl compounds (§2. V).

Van der Waals forces are electrostatic in nature. They are relatively weak forces (*i.e.*, in comparison with *bond forces*), but they are greater for compounds than for atoms and molecules of elements. In fact, the more asymmetrical the molecule, the

greater are the van der Waals forces. These forces originate from three different causes:

(i) Forces due to the interaction between the permanent dipole moments of the molecules (Keesom, 1916, 1921). These forces are known as **Keesom forces** or the **dipole-dipole effect**, and are dependent on temperature.

(ii) Forces which result from the interaction of a *permanent* dipole and *induced* dipoles. Although a molecule may not possess a permanent dipole, nevertheless a dipole may be induced under the influence of neighbouring molecules which do possess a permanent dipole (Debye, 1920, 1921). These forces are known as **Debye forces**, the **dipole-induced dipole effect** or **induction effect**, and are almost independent of temperature.

(iii) London (1930) showed from wave mechanics that a third form of van der Waals forces is also acting. A nucleus and its "electron cloud" are in a state of vibration, and when two atoms are sufficiently close to each other, the two nuclei and the two electron clouds tend to vibrate together, thereby leading to attraction between different molecules. These forces are known as **London forces, dispersion forces**, or the **wave-mechanical effect**, and are independent of temperature.

It should be noted that the induced forces are smaller than the other two, and that the dispersion forces are usually the greatest.

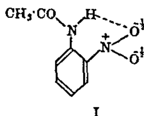
It can now be seen that all those physical properties which depend on intermolecular forces, *e.g.*, melting point, boiling point, viscosity, etc., will thus be largely determined by the van der Waals forces. Van der Waals forces are also responsible for the formation of "molecular compounds" such as "picrates" formed from picric acid and aromatic hydrocarbons or aromatic amines; these are "held together" by the Keesom and Debye forces.

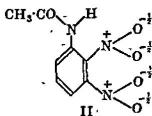
§3. The hydrogen bond. A particularly important case of electrostatic attraction is that which occurs in *hydrogen bonding* (Vol. I, Ch. II); it occurs mainly in compounds containing hydroxyl or imino groups. There are two types of hydrogen bonding, *intermolecular* and *intramolecular*. Intermolecular bonding gives rise to association, thereby raising the boiling point; it also raises the surface tension and the viscosity, but lowers the dielectric constant. Intermolecular hydrogen bonding may exist in compounds in the liquid or solid state, and its formation is very much affected by the shape of the molecules, *i.e.*, by the *spatial* or *steric* factor; *e.g.*, *n*-pentanol is completely associated, whereas *tert*-pentanol is only partially associated. Intermolecular hydrogen bonding is also responsible for the formation of various molecular compounds

(*cf* §2), and also affects solubility if the compound can form hydrogen bonds with the solvent.

Intramolecular hydrogen bonding gives rise to *chelation, i.e.*, ring formation, and this normally occurs only with the formation of five- or six-membered rings. Chelation has been used to explain the volatility of *ortho*-compounds such as *o*-halogenophenols and *o*-nitrophenols (as compared with the corresponding *m*- and *p*-derivatives). Chelation has also been used to account for various *ortho*-substituted benzoic acids being stronger acids than the corresponding *m*- and *p*-derivatives (see Vol. I, Ch. XXVIII).

When chelation occurs, the ring formed must be planar or almost planar. Should another group be present which prevents the formation of a *planar* chelate structure, then chelation will be diminished or even completely inhibited (Hunter *et al.*, 1938; *cf.* steric inhibition of resonance, Vol. I, Ch. XXVIII). Compound I is chelated, but II is associated and not chelated. In I the *o*-nitro-group can enter into the formation of a *planar* six-membered ring.





In II, owing to the strong repulsion between the negatively charged oxygen atoms of the two nitro-groups, the plane of each nitro-group will tend to be perpendicular to the plane of the benzene ring, and consequently a chelated *planar* six-membered ring cannot be formed.

The presence of hydrogen bonding may be detected by various means, *e.g.*, infra-red absorption spectra, X-ray analysis, electron diffraction, examination of boiling points, melting points, solubility, etc. The best method appears to be that of infra-red absorption spectra (see §15b).

§4. Melting point. In most solids the atoms or molecules are in a state of vibration about their fixed mean positions. These vibrations are due to the thermal energy and their amplitudes are small compared with interatomic distances. As the temperature of the solid is raised, the amplitude of vibration increases and a point is reached when the crystalline structure suddenly becomes unstable; this is the melting point.

In many homologous series the melting points of the *n*-members

rise continuously, tending towards a maximum value. On the other hand, some homologous series show an alternation or oscillation of melting points—"the saw-tooth rule", *e.g.*, in the fatty acid series the melting point of an "even" acid is higher than that of the "odd" acid immediately below and above it. It has been shown by X-ray analysis that this alternation of melting points depends on the packing of the crystals. The shape of the molecule is closely related to the melting point; the more symmetrical the molecule, the higher is the melting point. Thus with isomers, branching of the chain (which increases symmetry) usually raises the melting point; also *trans* isomers usually have a higher melting point than the *cis*, the former having greater symmetry than the latter (see §5. IV). In the benzene series, of the three disubstituted derivatives, the *p*-compound usually has the highest melting point.

Apart from the usual van der Waals forces which affect melting points, hydrogen bonding may also play a part, *e.g.*, the melting point of an alcohol is higher than that of its corresponding alkane. This may be attributed to hydrogen bonding, which is possible in the former but not in the latter.

Various *empirical* formulae have been developed from which it is possible to calculate melting points; these formulae, however, only relate members of an *homologous* series.

The method of mixed melting points has long been used to identify a compound, and is based on the principle that two different compounds mutually lower the melting point of each component in the mixture. This method, however, is unreliable when the two compounds form a solid solution.

§5. Boiling point. The boiling point of a liquid is that temperature at which the vapour pressure is equal to that of the external pressure. Thus the boiling point varies with the pressure, being raised as the pressure is increased.

In an homologous series, the boiling point usually increases regularly for the *n*-members, *e.g.*, Kopp (1842) found that with the aliphatic alcohols, acids, esters, etc., the boiling point is raised by 19° for each increase of CH_2 in the composition. In the case of isomers the greater the branching of the carbon chain, the lower is the boiling point. Calculation has shown that the boiling point of the *n*-alkanes should be proportional to the number of carbon atoms in the molecule. This relationship, however, is not observed in practice, and the cause of this deviation still remains to be elucidated. One strongly favoured theory attributes the cause to the fact that the carbon chains of *n*-alkanes in the liquid phase

exist largely in a coiled configuration. As the branching increases, the coil becomes denser, and this lowers the boiling point.

In aromatic disubstituted compounds the boiling point of the *ortho*-isomer is greater than that of the *meta*-isomer which, in turn, may have a higher boiling point than the *para*-isomer, but in many cases the boiling points are about the same.

Since the boiling point depends on the van der Waals forces, any structural change which affects these forces will consequently change the boiling point. One such structural change is the branching of the carbon chain (see above). Another type of change is that of substituting hydrogen by a negative group. This introduces a dipole moment (or increases the value of an existing dipole moment), thereby increasing the attractive forces between the molecules and consequently raising the boiling point, *e.g.*, the boiling points of the nitro-alkanes are very much higher than those of the corresponding alkanes. *Cis* compounds usually have higher boiling points than the corresponding *trans* isomers, the reason again being attributed to the larger dipole moments of the former. This "regularity" has been used to differentiate between geometrical isomers (see §5. IV). The possibility of intermolecular hydrogen bonding also raises the boiling point, *e.g.*, alcohols boil at higher temperatures than the corresponding alkanes.

§6. Solubility. It is believed that solubility depends on the following intermolecular forces: solvent/solute; solute/solute; solvent/solvent. The solubility of a non-electrolyte in water depends, to a very large extent, on whether the compound can form hydrogen bonds with the water, *e.g.*, the alkanes are insoluble, or almost insoluble, in water. Methane, however, is more soluble than any of its homologues. The reason for this is uncertain; hydrogen bonding with water is unlikely, and so other factors must play a part, *e.g.*, molecular size. A useful guide in organic chemistry is that "like dissolves like", *e.g.*, if a compound contains a hydroxyl group, then the best solvents for that compound also usually contain hydroxyl groups (hydrogen bonding between solvent and solute is possible). This "rule" is accepted by many who use the word "like" to mean that the cohesion forces in both solvent and solute arise from the same source, *e.g.*, alkanes and alkyl halides are miscible; the cohesion forces of both of these groups of compounds are largely due to dispersion forces.

In some cases solubility may be due, at least partly, to the formation of a compound between the solute and the solvent, *e.g.*, ether dissolves in concentrated sulphuric acid with the formation of an oxonium salt, $\{(C_2H_5)_2OH\}^+HSO_4^-$.

§7. Viscosity. Viscosity (the resistance to flow due to the internal friction in a liquid) depends, among other factors, on the van der Waals forces acting between the molecules. Since these forces depend on the shape and size of the molecules, the viscosity will also depend on these properties. At the same time, since the Keesom forces (§2) depend on temperature, viscosity will also depend on temperature; other factors, however, also play a part.

A number of relationships have been found between the viscosity of pure liquids and their chemical structure, *e.g.*,

(i) In an homologous series, viscosity increases with the molecular weight.

(ii) With isomers the viscosity of the *n*-compound is greater than that of isomers with branched carbon chains.

(iii) Abnormal viscosities are shown by *associated* liquids. Viscosity measurements have thus been used to determine the degree of association in liquids.

(iv) The viscosity of a *trans* compound is greater than that of the corresponding *cis* isomer.

Equations have been developed relating viscosity to the shape and size of *large* molecules (*macromolecules*) in solution, and so viscosity measurements have offered a means of determining the shape of, *e.g.*, proteins, and the molecular weight of, *e.g.*, polysaccharides.

§8. Molecular volumes. The molecular volume of a liquid in millilitres (V_m) is given by the equation

$$V_m = \frac{\text{gram molecular weight}}{\text{density}}$$

The relation between molecular volume and chemical composition was studied by Kopp (1839–1855). Since the density of a liquid varies with the temperature, it was necessary to choose a standard temperature for comparison. Kopp chose the boiling point of the liquid as the standard temperature. This choice was accidental, but proved to be a fortunate one since the absolute boiling point of a liquid at atmospheric pressure is approximately two-thirds of the critical temperature, *i.e.*, Kopp unknowingly compared liquids in their corresponding states, the theory of which did not ppear until 1879. As a result of his work, Kopp was able to mpile a table of atomic volumes based on the assumption that e molecular volume was an additive property, *e.g.*,

C	11·0	Cl	22·8
H	5·5	Br	27·8
O (C=O)	12·2	I	37·5
O(OH)	7·8		

It should be noted that Kopp found that the atomic volume of oxygen (and sulphur) depended on its state of combination. Kopp also showed that the molecular volume of a compound can be calculated from the sum of the atomic volumes, *e.g.*, acetone, $CH_3 \cdot CO \cdot CH_3$.

3C	= 33·0	Molecular weight of acetone = 58
6H	= 33·0	Density at b.p. = 0·749
O(CO)	= 12·2	$\therefore$ molecular volume (*obs.*) $= \frac{58}{0 \cdot 749} = 77 \cdot 4$
	78·2 (*calc.*)	

Further work has shown that the molecular volume is not strictly additive, but also partly constitutive (as recognised by Kopp who, however, tended to overlook this feature). If purely additive, then isomers with *similar* structures will have the same molecular volume. This has been found to be the case for, *e.g.*, isomeric esters, but when the isomers belong to different homologous series, the agreement may be poor.

Later tables have been compiled for atomic volumes with structural corrections. Even so, the relation breaks down in the case of highly polar liquids where the attractive forces between the molecules are so great that the additive (and structural) properties of the atomic volumes are completely masked.

§9. **Parachor.** Macleod (1923) introduced the following equation:

$$\gamma = C(d_l - d_v)^4$$

where γ is the surface tension, d_l and d_v the densities of the liquid and vapour respectively, and C is a constant which is independent of the temperature.

Macleod's equation can be rewritten as:

$$\frac{\gamma^{\frac{1}{4}}}{d_l - d_v} = C^{\frac{1}{4}}$$

Sugden (1924) multiplied both sides of this equation by the molecular weight, M, and pointed out that the expression

$$\frac{M\gamma^{\frac{1}{4}}}{d_l - d_v} = MC^{\frac{1}{4}} = [P]$$

should also be valid. Sugden called the constant P for a given compound the *parachor* of that compound. Provided the temperature is not too high, d_v will be negligible compared with d_l, and so we have

$$[P] = \frac{M\gamma^{\frac{1}{4}}}{d_l}$$

Hence the parachor represents the molecular volume of a liquid at the temperature when its surface tension is unity. Thus a comparison of parachors of different liquids gives a comparison of molecular volumes at temperatures at which liquids have the same surface tension. By this means allowance is made for the van der Waals forces, and consequently the comparison of molecular volumes is carried out under comparable conditions.

The parachor is largely an additive property, but it is also partly constitutive. The following table of atomic and structural parachors is that given by Mumford and Phillips (1929).

C	9·2	Single bond	0
H	15·4	Co-ordinate bond	0
O	20	Double bond	19
N	17·5	Triple bond	38
Cl	55	3-Membered ring	12·5
Br	69	4- ,, ,,	6
I	90	5- ,, ,,	3
S	50	6- ,, ,,	0·8
		7- ,, ,,	− 4

The parachor has been used to enable a choice to be made between alternative structures, *e.g.*, structures I and II had been suggested for *p*-benzoquinone. Most of the chemical evidence


I


II

favoured I, but Graebe (1867) proposed II to explain some of the properties of this compound (see Vol. I). The parachor has been used to decide between these two:

[P] calculated for I is 233·6;
$[6 \times 9{\cdot}2 + 4 \times 15{\cdot}4 + 2 \times 20 + 4 \times 19 + 0{\cdot}8]$

[P] calculated for II is 215·4;
$[6 \times 9{\cdot}2 + 4 \times 15{\cdot}4 + 2 \times 20 + 3 \times 19 + 2 \times 0{\cdot}8]$

[P] observed is 236·8. This indicates structure I.

According to Sutton (1952), the parachor is not a satisfactory property for the analysis of molecular structure. It is, however, still useful as a physical characteristic of the liquid-vapour system.

§10. Refrachor. Joshi and Tuli (1951) have introduced a new physical constant which they have named the *refrachor*, [F]. This

has been obtained by associating the parachor, [P], with the refractive index, (n_D^{20}), according to the following equation:

$$[F] = -[P] \log (n_D^{20} - 1)$$

The authors have found that the observed refrachor of any compound is composed of two constants, one dependent on the nature of the atoms, and the other on structural factors, *e.g.*, type of bond, size of ring, etc., *i.e.*, the refrachor is partly additive and partly constitutive. Joshi and Tuli have used the refrachor to determine the percentage of tautomers in equilibrium mixtures, *e.g.*, they found that ethyl acetoacetate contains 7·7% enol, and penta-2:4-dione 72·4% enol.

§11. Refractive index. Lorentz and Lorenz (1880) simultaneously showed that

$$R = \frac{n^2 - 1}{n^2 - 2} \cdot \frac{M}{d}$$

where R is the *molecular refractivity*, n the refractive index, M the molecular weight, and d the density. The value of n depends on the wavelength and on temperature; d depends on temperature.

Molecular refractivity has been shown to have both additive and constitutive properties. The following table of atomic and structural refractivities has been calculated for the H_α line.

C	2·413	Cl	5·933
H	1·092	Br	8·803
O(OH)	1·522	I	13·757
O(CO)	2·189	Double bond (C=C)	1·686
O(ethers)	1·639	Triple bond (C≡C)	2·328

Molecular refractivities have been used to determine the structure of compounds, *e.g.*, terpenes (see §25. VIII). They have also been used to detect the presence of tautomers and to calculate the amount of each form present. Let us consider ethyl acetoacetate as an example; this behaves as the keto form $CH_3{\cdot}CO{\cdot}CH_2{\cdot}CO_2C_2H_5$, and as the enol form $CH_3{\cdot}C(OH){=}CH{\cdot}CO_2C_2H_5$. The calculated molecular refractivities of these forms are:

$CH_3{\cdot}CO{\cdot}CH_2{\cdot}CO_2C_2H_5$		$CH_3{\cdot}C(OH) = CH{\cdot}CO_2C_2H_5$	
6 C	= 14·478	6 C	= 14·478
10 H	= 10·92	10 H	= 10·92
2 O (CO)	= 4·378	O (OH)	= 1·522
O (ether)	= 1·639	O (CO)	= 2·189
		O (ether)	= 1·639
	31·415	Double bond	= 1·686
			32·434

with rise in temperature, but many cases are known where the rotatory power decreases. Hence, for comparison, it is necessary to state the temperature; in practice, measurements are usually carried out at 20 or 25°.

(iv) *The solvent.* The nature of the solvent affects the rotation, and so it is necessary to state the solvent used in the measurement of the rotatory power. There appears to be some relation between the effect of a solvent on rotatory power and its dipole moment.

(v) *The concentration.* The rotation appears to be independent of the concentration provided that the solution is dilute. In concentrated solutions, however, the rotation varies with the concentration, the causes for this have been attributed to association, dissociation, or solvation (see also vi).

(vi) The amount of rotation exhibited by a given substance when all the preceding factors (i–v) have been fixed may be varied by the presence of other compounds which are not, in themselves, optically active, *e.g.*, inorganic salts. It is important to note in this connection that optically active acids or bases, in the form of their salts, give rotations which are independent of the nature of the non-optically active ion *provided that the solutions are very dilute.* In very dilute solutions, salts are completely dissociated, and it is only the optically active ion which then contributes to the rotation. The rotation of a salt formed from an optically active acid and an optically active base reaches a constant value in dilute solutions, and the rotation is the sum of the rotations of the anion and cation. This property has been used to detect optical activity (see §5a. VI).

When recording the rotations of substances, the value commonly given is the **specific rotation**, $\left[\alpha\right]_\lambda^t$. This is obtained from the equation:

$$\left[\alpha\right]_\lambda^t = \frac{\alpha_\lambda}{l \times d} \quad or \quad \left[\alpha\right]_\lambda^t = \frac{\alpha_\lambda^t}{l \times c}$$

where l is the thickness of the layer in decimetres, d the density of the liquid (if it is a pure compound), c the number of grams of substance per millilitre of solution (if a solution is being examined), α the *observed* rotation, t the temperature, and λ the wavelength of the light used. The solvent should also be stated (see iv).

The **molecular rotation**, $\left[M\right]_\lambda^t$, is obtained by multiplying the specific rotation by the molecular weight, M. Since large numbers are usually obtained, a common practice is to divide the result by one hundred; thus:

$$[M]_{\lambda}^{t} = \frac{[\alpha]_{\lambda}^{t} \times M}{100}$$

The relation between structure and optical activity is discussed later (see §§2, 3. II). The property of optical activity has been used in the study of the configuration of molecules and mechanisms of various reactions, and also to decide between alternative structures for a given compound. The use of optical rotations in the determination of structure depends largely on the application of two rules.

(i) **Rule of Optical Superposition** (van't Hoff, 1894): When a compound contains two or more asymmetric centres, the total rotatory power of the molecules is the algebraic sum of the contributions of each asymmetric centre. This rule is based on the assumption that the contribution of each asymmetric centre is independent of the other asymmetric centres present. It has been found, however, that the contribution of a given asymmetric centre is affected by neighbouring centres and also by the presence of chain-branching and unsaturation. Hence the rule, although useful, must be treated with reserve (see also §6. VII).

A more satisfactory rule is the **Rule of Shift** (Freudenberg, 1933): If two asymmetric molecules A and B are changed in the same way to give A′ and B′, then the differences in molecular rotation (A′ − A) and (B′ − B) are of the same sign (see, *e.g.*, §4b. XI).

(ii) **Distance Rule** (Tschugaev, 1898): The effect of a given structural change on the contribution of an asymmetric centre decreases the further the centre of change is from the asymmetric centre.

Only asymmetric molecules have the power, under normal conditions, to rotate the plane of polarisation (of plane-polarised light). Faraday (1845), however, found that any transparent substance can rotate the plane of polarisation when placed in a strong magnetic field. This property of **magnetic optical rotation (Faraday effect)** is mainly an additive one, but is also partly constitutive.

§13. Dipole moments. When the centres of gravity of the electrons and nuclei in a molecule do not coincide, the molecule will possess a *permanent* dipole moment, μ, the value of which is given by $\mu = e \times d$, where e is the electronic charge, and d the distance between the charges (positive and negative centres). Since e is of the order of 10^{-10} e.s.u., and d 10^{-8} cm., μ is therefore of the order 10^{-18} e.s.u. This unit is known as the Debye (D), in honour of Debye, who did a great deal of work on dipole moments. The dipole moment is a vector quantity, and its direction in a molecule is often indicated by an arrow parallel to the line joining

the points of charge, and pointing towards the negative end, *e.g.*, $\overset{+\!\!\longrightarrow}{\text{H—Cl}}$ (Sidgwick, 1930). The greater the value of the dipole moment, the greater is the *polarity* of the bond. It should be noted that the terms *polar* and *non-polar* are used to describe bonds, molecules and groups. Bond dipoles are produced because of the different electron-attracting powers of atoms (or groups) joined by that bond. This unequal electronegativity producing a dipole moment seems to be a satisfactory explanation for many simple molecules, but is unsatisfactory in other cases. Thus a number of factors must operate in determining the value of the dipole moment. It is now believed that *four* factors contribute to the bond moment:

(i) The unequal sharing of the bonding electrons arising from the different electronegativities of the two atoms produces a dipole moment.

(ii) In covalent bonds a dipole is produced because of the difference in size of the two atoms. The centres of gravity (of the charges) are at the nucleus of each contributing atom. Thus, if the atoms are different in size, the resultant centre of gravity is not at the mid-point of the bond, and so a bond moment results.

(iii) Hybridisation of orbitals produces asymmetric atomic orbitals; consequently the centres of gravity of the hybridised orbitals are no longer at the parent nuclei. Only if the orbitals are pure *s*, *p* or *d*, are the centres of gravity at the parent nuclei. Thus hybridised orbitals produce a bond moment.

(iv) Lone-pair electrons (*e.g.*, on the oxygen atom in water) are not "pure" *s* electrons; they are "impure" because of hybridisation with *p* electrons. If lone-pair electrons were not hybridised, their centre of gravity would be at the nucleus; hybridisation, however, displaces the centre of gravity from the nucleus and so the asymmetric orbital produced gives rise to a bond moment which may be so large as to outweigh the contributions of the other factors to the dipole moment.

The following points are useful in organic chemistry:

(i) In the bond H—Z, where Z is any atom other than hydrogen or carbon, the hydrogen atom is the positive end of the dipole, *i.e.*, $\overset{+\!\!\longrightarrow}{\text{H—Z}}$.

(ii) In the bond C—Z, where Z is any atom other than carbon, the carbon atom is the positive end of the dipole, *i.e.*, $\overset{+\!\!\longrightarrow}{\text{C—Z}}$ (Coulson, 1942).

(iii) When a molecule contains two or more polar bonds, the resultant dipole moment of the molecule is obtained by the vectorial addition of the constituent bond dipole moments. A sym-

metrical molecule will thus be non-polar, although it may contain polar bonds, *e.g.*, CCl_4 has a zero dipole moment although each C—Cl bond is strongly polar.

When any molecule (polar or non-polar) is placed in an electric field, the electrons are displaced from their normal positions (towards the positive pole of the external field). The positive nuclei are also displaced (towards the negative pole of the external field), but their displacement is much less than that of the electrons because of their relatively large masses. These displacements give rise to an *induced* dipole, and this exists only while the external electric field is present. The value of the induced dipole depends on the strength of the external field and on the *polarisability* of the molecule, *i.e.*, the ease with which the charged centres are displaced by the external field. If P is the total dipole moment, P_μ the permanent dipole moment, and P_α the induced dipole moment, then

$$P = P_\mu + P_\alpha$$

P_μ decreases as the temperature rises, but P_α is independent of the temperature. The value of P *in solution* depends on the nature of the solvent and on the concentration.

By means of dipole moment measurements, it has been possible to get a great deal of information about molecules, *e.g.*,

(i) Configurations of molecules have been ascertained, *e.g.*, water has a dipole moment and hence the molecule cannot be linear. In a similar way it has been shown that ammonia and phosphorus trichloride are not flat molecules.

(ii) Orientations in benzene derivatives have been examined by dipole moments (see Vol. I). At the same time, this method has shown that the benzene molecule is flat.

(iii) Dipole moment measurements have been used to distinguish between geometrical isomers (see §5. IV).

(iv) Dipole moments have been used to demonstrate the existence of resonance and to elucidate electronic structures.

(v) Energy differences between different conformations (see §4a. II) have been calculated from dipole moment data.

(vi) The existence of dipole moments gives rise to association, the formation of molecular compounds, etc.

§14. Magnetic susceptibility. When a substance is placed in a magnetic field, the substance may or may not become magnetised. If I is the *intensity of magnetisation* induced, and H the strength of the magnetic field inducing it, then the **magnetic susceptibility**, κ, is given by

$$\kappa = \frac{I}{H}$$

The *magnetic induction*, B, is given by

$$B = H + 4\pi I$$

$$\text{Since } I = \kappa H,\ B = H(1 + 4\pi\kappa)$$

The quantity $1 + 4\pi\kappa$ is called the *magnetic permeability*, μ.

Elements other than iron, nickel and cobalt (which are *ferromagnetic*) may be divided into two groups:

(i) **Paramagnetic**: in this group μ is greater than unity and κ is therefore positive.

(ii) **Diamagnetic**: in this group μ is less than unity and κ is therefore negative.

All compounds are either paramagnetic or diamagnetic. Paramagnetic substances possess a permanent magnetic moment and consequently orient themselves along the external magnetic field. Diamagnetic substances do not possess a permanent magnetic moment, and tend to orient themselves at right angles to the external magnetic field.

Electrons, because of their spin, possess magnetic dipoles. When electrons are paired (*i.e.*, their spins are anti-parallel), then the magnetic field is cancelled out. Most organic compounds are diamagnetic, since their electrons are paired. "Odd electron molecules", however, are paramagnetic.

Magnetic susceptibility has been used to obtain information on the nature of bonds and the configuration of co-ordination compounds. Organic compounds which are paramagnetic are generally free radicals (odd electron molecules), and the degree of dissociation of, *e.g.*, hexaphenylethane into triphenylmethyl has been measured by means of its magnetic susceptibility.

§15. **Absorption spectra.** When light (this term will be used for electromagnetic waves of any wavelength) is absorbed by a molecule, the molecule undergoes transition from a state of lower to a state of higher energy. If the molecule is monatomic, the energy absorbed can only be used to raise the energy levels of electrons. If, however, the molecule consists of more than one atom, the light absorbed may bring about changes in electronic, rotational or vibrational energy. Electronic transitions give absorption (or emission) in the visible and ultra-violet parts of the spectrum, whereas rotational and vibrational changes give absorption (or emission) respectively in the far and near infra-red. Electronic transitions may be accompanied by the other two. A study of these energy changes gives information on the structure of molecules.

Spectrum	Wavelength (A)
Ultra-violet	2,000–4,000
Visible	4,000–7,500
Near infra-red	7,500–15 × 10⁴
Far infra-red	15 × 10⁴–100 × 10⁴

The position of the absorption band can be given as the wavelength λ (cm., μ, A, mμ) or as the wave number, $\tilde{\nu}$ (cm.$^{-1}$).

1 μ (micron) = 10^{-3} mm. 1 mμ (millimicron) = 10^{-6} mm.
1 A. (Angstrom) = 10^{-8} cm. = 10^{-7} mm. 1 mμ = 10 A.

$$\lambda\ (\mu) = \frac{10^4}{\tilde{\nu}\ (\text{cm.}^{-1})}$$

$$\tilde{\nu}\ (\text{cm.}^{-1}) = \frac{1}{\lambda\ (\text{cm.})} = \frac{10^4}{\lambda\ (\mu)} = \frac{10^8}{\lambda\ (\text{A})}$$

If I_0 is the intensity of an incident beam of monochromatic light, and I that of the emergent beam which has passed through an absorbing medium of thickness l, then

$$I = I_0 10^{-\varepsilon l} \quad or \quad \log_{10}\frac{I_0}{I} = \varepsilon l$$

where ε is the *extinction coefficient* of the medium. The ratio I_0/I is called the *transmittance* of the medium, and the reciprocal the *opacity*; the function $\log_{10} I_0/I$ is called the *density* (*d*).

If the absorbing substance is in solution (the solvent being *colourless*), and if c is the concentration (number of grams per litre), then

$$I = I_0 10^{-\varepsilon c l}$$

This equation is **Beer's law** (1852), and is obeyed by most solutions provided they are *dilute*. In more concentrated solutions there may be divergencies from Beer's law, and these may be caused by association, changes in solvation, etc.

If the extinction coefficient is plotted against the wavelength of the light used, the *absorption curve* of the compound is obtained, and this is characteristic for a *pure* compound (under identical conditions).

§15a. Ultra-violet and visible absorption spectra. When a molecule absorbs light; it will be raised from the ground state to an excited state. The position of the absorption band depends on the difference between the energy levels of the ground and excited states. Any change in the structure of the molecule which alters the energy difference between the ground and excited states will thus affect the position of the absorption band. This shifting of

bands (in the ultra-violet and visible regions) is concerned with the problem of colour (see Vol. I, Ch. XXXI).

With *few exceptions, only molecules containing multiple bonds* give rise to absorption in the near ultra-violet. In compounds containing only one multiple-bond group, the intensity of the absorption maxima may be very low, but when several of these groups are present in conjugation, the absorption is strong, *e.g.*, an isolated oxo (carbonyl) group has an absorption at λ_{max} 2750 A; an isolated ethylene bond has an absorption at $\lambda_{max.}$ 1950 A. When a compound contains an oxo group conjugated with an ethylenic bond, *i.e.*, the compound is an $\alpha\beta$-unsaturated oxo compound, the two bands no longer occur in their original positions, but are shifted to 3100–3300 A and 2200–2600 A, respectively. Thus, in a compound *in which the presence of an ethylenic bond and an oxo* group has been demonstrated (by chemical methods), it is also *possible to tell, by examination of the ultra-violet absorption* spectrum, whether the two groups are conjugated or not (see, *e.g.*, cholestenone, §3(ii). XI).

Ultra-violet and visible absorption spectra have also been used to differentiate between geometrical isomers and to detect the presence or absence of restricted rotation in diphenyl compounds (§2. V).

§15b. Infra-red spectra. In a molecule which has some [illegible] atoms vibrate with frequencies [illegible] ing forces [illegible] its equilibrium configuration. The energy for these vibrations is absorbed *from the incident light, and thereby gives rise to a vibrational* spectrum. A given bond has a characteristic absorption band, but the frequency depends, *to some extent, on the nature of the* other atoms joined to the two atoms under consideration. It is thus possible to ascertain the nature *of bonds (and therefore groups)* in unknown compounds by comparing their infra-red spectra with tables of infra-red absorption spectra. At the same time *it is also* possible to verify tentative structures (obtained from chemical evidence) by comparison with spectra of *similar* compounds of known structure.

The study of infra-red spectra leads to information on many types of problems, *e.g.*,

(i) Infra-red spectroscopy has been used to distinguish between geometrical isomers, and recently Kuhn (1950) has shown that the spectra of the stereoisomers methyl α- and β-glycosides are different. It also appears that enantiomorphs in the *solid* phase often exhibit

different absorption spectra. Infra-red spectroscopy has also been a very valuable method in conformational studies (see §11. IV).

(ii) The three isomeric disubstituted benzenes have characteristic absorption bands, and this offers a means of determining their orientation.

(iii) Infra-red spectroscopy has given a great deal of information about the problem of free rotation about a single bond; *e.g.*, since the intensity of absorption is proportional to the concentration, it has been possible to ascertain the presence and amounts of different conformations in a mixture (the intensities vary with the temperature when two or more conformations are present).

(iv) Tautomeric mixtures have been examined and the amounts of the tautomers obtained. In many cases the *existence* of tautomerism can be ascertained by infra-red spectroscopy (*cf.* iii).

(v) Infra-red spectroscopy appears to be the best means of ascertaining the presence of hydrogen bonding (both in association and chelation). In "ordinary" experiments it is not possible to distinguish between intra- and intermolecular hydrogen bonding. These two modes of bonding can, however, be differentiated by obtaining a series of spectra at different dilutions. As the dilution increases, the absorption due to intermolecular hydrogen bonding decreases, whereas the intramolecular hydrogen-bonding absorption is unaffected.

(vi) It is possible to evaluate dipole moments from infra-red spectra.

(vii) When a bond between two atoms is stretched, a restoring force immediately operates. If the distortion is *small*, the restoring force may be assumed to be directly proportional to the distortion, *i.e.*,

$$f \propto d \quad \text{or} \quad f = kd$$

where k is the *stretching force constant* of the bond. It is possible to calculate the values of these force constants from infra-red (vibrational) spectra.

(viii) The far infra-red or micro-wave region contains the *pure rotational* spectrum. Micro-wave spectroscopy (a recent development) offers a very good method for measuring bond lengths. It is possible to calculate atomic radii from bond-lengths, but the value depends on whether the bond is single, double or triple, and also on the charges (if any) on the atoms concerned. Thus the character of a bond can be ascertained from its length, *e.g.*, if a bond length (determined experimentally) differs significantly from the sum of the atomic radii, then the bond is not "normal". Resonance may be the cause of this.

Some atomic covalent radii (in Angstroms) are:

H	0·30	N (single)	0·70	Cl	0·99
C (single)	0·77	N (double)	0·61	Br	1·14
C (double)	0·67	N (triple)	0·55	I	1·33
C (triple)	0·60	O (single)	0·66	S	1·04
		O (double)	0·57		

Micro-wave spectroscopy is particularly useful for information on the molecular structure of polar gases, and is also used for showing the presence of free radicals.

§15c. Raman spectra. When a beam of monochromatic light passes through a transparent medium, most of the light is transmitted or scattered without change in wavelength. Some of the light, however, is converted into *longer* wavelengths, *i.e.*, *lower* frequency (a smaller amount of the light may be changed into shorter wavelengths, *i.e.*, higher frequency). The change from *higher to lower* frequency is known as the **Raman effect** (**Raman shift**). It is independent of the frequency of the light used, but is characteristic for a given bond.

Raman spectra have been used to obtain information on struc- [illegible] tion, and to give information on force constants of bonds.

§16. X-ray analysis. X-rays may be used with gases, liquids or solids, but in organic chemistry they are usually confined to solids, which may be single crystals, or substances consisting of a mass of minute crystals (*powder method*), or fibres. When X-rays (wavelength 0·7 to 1·5 A) fall on solids, they are diffracted to produce patterns (formed on a photographic film). Since X-rays are diffracted mainly by the orbital electrons of the atoms, the diffraction will be a function of the atomic number. Because of this, it is difficult to differentiate between atoms whose atomic numbers are very close together, *e.g.*, carbon and nitrogen. Furthermore, since the scattering power of hydrogen atoms (for X-rays) is very low, it is normally impossible to locate these atoms except in very favourable conditions, and then only with fairly simple compounds

Two problems are involved in the interpretation of X-ray diffraction patterns, *viz.*, the dimensions of the unit cell and the positions of the individual atoms in the molecule. The positions of the

diffracted beams depend on the dimensions of the unit cell. A knowledge of these dimensions leads to the following applications :

(i) Identification of substances ; this is done by looking up tables of unit cells.

(ii) Determination of molecular weights. If V is the volume of the unit cell, d the density of the compound, and n the number of molecules in a unit cell, then the molecular weight, M, is given by

$$M = \frac{Vd}{n}$$

(iii) Determination of the shapes of molecules. Many long-chain polymers exist as fibres, *e.g.*, cellulose, keratin. These fibres are composed of bundles of tiny crystals with one axis parallel, or nearly parallel, to the fibre axis. When X-rays fall on the fibre in a direction perpendicular to its length, then the pattern obtained is similar to that from a single crystal rotated about a principal axis. It is thus possible to obtain the unit cell dimensions of such fibres (see, *e.g.*, rubber, §33. VIII).

The intensities of the diffracted beams depend on the positions of the atoms in the unit cell. A knowledge of these relative intensities leads to the following applications :

(i) Determination of bond lengths, valency angles, and the general electron distribution in molecules.

(ii) Determination of molecular symmetry. This offers a means of distinguishing between geometrical isomers, and also of ascertaining the shape of a molecule, *e.g.*, the diphenyl molecule has a centre of symmetry, and therefore the two benzene rings must be coplanar (see §2. V).

(iii) Determination of structure. This application was originally used for compounds of *known* structure. Trial models based on the structure of the molecule were compared with the X-ray patterns, and if they "fitted", *confirmed* the structure already accepted. If the patterns did not fit, then it was necessary to look for another structural formula. More recently, however, X-ray analysis has been applied to compounds of unknown or partially known structures, *e.g.*, penicillin (§6a. XVIII).

(iv) X-ray analysis has been used to elucidate the conformations of rotational isomers (§4a. II), and also to determine the *absolute* configurations of enantiomorphs (§5. II).

§17. Electron diffraction. Electron diffraction is another direct method for determining the spatial arrangement of atoms in a molecule, and is usually confined to gases or compounds in the vapour state, but may be used for solids and liquids. Electrons exhibit a dual behaviour, particle or wave, according to the nature

of the experiment. The wavelength of electrons is inversely proportional to their momentum: the wavelength is about 0·06 A for the voltages generally used. Because of their small diffracting power, hydrogen atoms are difficult, if not impossible, to locate.

By means of electron diffraction it is possible to obtain values of bond lengths and the size and shape of molecules, particularly macromolecules. Electron diffraction studies have been particularly useful in the investigation of conformations in *cyclo*hexane compounds (see §11. IV).

§18. Neutron crystallography. A beam of *slow* neutrons is diffracted by crystalline substances. The equivalent wavelength of a slow beam of neutrons is 1 A, and since this is of the order of interatomic distances in crystals, the neutrons will be diffracted. This method of analysis is particularly useful for determining the positions of *light* atoms, a problem which is very difficult, and often impossible, with X-ray analysis. Thus neutron diffraction is extremely useful for locating hydrogen atoms.

In addition to studying solids, neutron diffraction has also been applied to gases, pure liquids and solutions.

§19. Nuclear magnetic resonance. In addition to magnetic moments due to electrons (§14), there are also magnetic moments due to nuclei. Nuclei contain protons and neutrons which spin about their own axes and hence these particles possess magnetic moments. In most nuclei the spins are not cancelled out and consequently such nuclei possess a resultant nuclear magnetic moment. Transitions among the various nuclear magnetic energy levels may be induced by means of an external oscillating magnetic field, and under certain conditions, resonance absorption occurs. To produce nuclear magnetic resonance, the alternating field must have a comparatively low frequency (radiofrequencies), and furthermore, the frequency of the absorbed energy can be varied by changing the strength of the applied field. The resonance frequencies of most magnetic nuclei lie between 0·1 and 40 megacycles for fields varying from 1,000 to 10,000 gauss.

Nuclear magnetic resonance has been used to provide information on molecular structure, molecular electron distribution, and hindered rotation. It has also been used for the measurement of keto-enol equilibria and for the detection of association, etc.

9716

READING REFERENCES

Partington, *An Advanced Treatise on Physical Chemistry*, Longmans Green. Vol. I–V (1949–1954).
Ferguson, *Electronic Structures of Organic Molecules*, Prentice-Hall (1952).
Ketelaar, *Chemical Constitution*, Elsevier (1953).
Gilman, *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). (i) Vol. II Ch. 23. Constitution and Physical Properties of Organic Compounds. (ii) Vol. III (1953). Ch. 2. Applications of Infra-red and Ultra-violet Spectra to Organic Chemistry.
Wells, *Structural Inorganic Chemistry*, Oxford Press (1950, 2nd ed.).
Syrkin and Dyatkina, *Structure of Molecules and the Chemical Bond*, Butterworth (1950; translated and revised by Partridge and Jordan).
Weissberger (Ed.), *Technique of Organic Chemistry*, Interscience Publishers. Vol. I (1949, 2nd ed.). Physical Methods of Organic Chemistry.
Berl (Ed.), *Physical Methods in Chemical Analysis*, Academic Press. Vol. I (1950); Vol. II (1951).
Sugden, *The Parachor and Valency*, Routledge (1930)
Waters, *Physical Aspects of Organic Chemistry*, Routledge and Kegan Paul (1950, 4th ed.).
Reilly and Rae, *Physico-Chemical Methods*, Methuen (Vol I and II, 1954, 5th ed.).
Bunn, *Chemical Crystallography*, Oxford Press (1946).
Robertson, *Organic Crystals and Molecules*, Cornell University Press (1953).
Ann. Reports (Chem. Soc.). (i) 1946, **43**, 5. Physical Aspects of the Hydrogen Bond. (ii) 1946, **43**, 141. The Hydrogen Bond (iii) 1946, **43**, 30. The Molecular Weight and Dimensions of Macromolecules in Solution. (iv) 1950, **47**, 85. Solutions of High Polymers. (v) 1950, **47**, 7; 1951, **48**, 16; 1954, **51**, 16. Molecular Structure. (vi) 1953, **50**, 9; 1954, **51**, 7; 1955, **52**, 72, 81. Molecular Spectra. (vii) 1950, **47**, 420. Neutron Crystallography.
Hartley, State of Solution of Colloidal Electrolytes, *Quart. Reviews (Chem. Soc.)*, 1948, **2**, 152.
Whiffen, Rotation Spectra, *Quart. Reviews (Chem. Soc.)*, 1950, **4**, 131.
Ubbelohde, Melting and Crystal Structure, *Quart. Reviews (Chem. Soc.)*, 1950, **4**, 356.
Jeffrey and Cruikshank, Molecular Structure Determination by X-Ray Crystal Analysis: Modern Methods and their Accuracy, *Quart. Reviews (Chem. Soc.)*, 1953, **7**, 335.
Rowlinson, Intermolecular Forces and Some Properties of Matter, *Quart. Reviews (Chem. Soc.)*, 1954, **8**, 168.
Quayle, The Parachors of Organic Compounds, *Chem. Reviews*, 1953, **53**, 439.
Joshi and Tuli, Refrachor: A New Physical Constant, *J.C.S.*, 1951, 837.
Herzberg, *Infra-red and Raman Spectra*, Van Nostrand Company (1945).
Stuart, *Die Struktur Des Freien Moleküls*, Springer-Verlag (1952).
Mizushima, *Structure of Molecules and Internal Rotation*, Academic Press (1954).

Ingold, *Structure and Mechanism in Organic Chemistry*, Bell and Sons (1953). Ch. III, Physical Properties of Molecules.
Ferguson, Hydrogen Bonding and Physical Properties of Substances, *J Chem Educ*, 1956, **33**, 267.
Giles *et al.*, Studies in Hydrogen-bond Formation, *J.C.S.*, **1956**, 559
Braude and Nachod (Ed.), *Determination of Organic Structures by Physical Methods*, Academic Press (1955).
Smith, *Electric Dipole Moments*, Butterworth (1955).
Richards, , in Crystals, *Quart. Reviews* (*Chem*
Ann. Revi ; —).
Newman (Ed), *Steric Effects in Organic Chemistry*, Wiley (1956). Ch 11. Steric Effects on Certain Physical Properties.
Smith, Nuclear Magnetic Resonance Absorption, *Quart. Reviews* (*Chem. Soc*), 1953, **7**, 279.
Wertz, Nuclear and Spin Magnetic Resonance, *Chem. Reviews*, 1955, **55**, 829.
Bellamy, *The Infra-red Spectra of Complex Molecules*, Methuen (1958, 2nd ed).

CHAPTER II

OPTICAL ISOMERISM

§1. Stereoisomerism. **Stereochemistry** is the "chemistry of space", *i.e.*, stereochemistry deals with the *spatial* arrangements of atoms and groups in a molecule. **Stereoisomerism** is exhibited by isomers having the *same* structure but differing in their spatial arrangement, *i.e.*, having different *configurations*. Different configurations are possible because carbon forms mainly covalent bonds and these have direction in space. The covalent bond is formed by the overlapping of atomic orbitals, the bond energy being greater the greater the overlap of the component orbitals. To get the maximum overlap of orbitals, the orbitals should be in the same plane. Thus *non-spherical* orbitals tend to form bonds in the direction of the greatest concentration of the orbital, and this consequently produces a *directional* bond (see also Vol. I, Ch. II).

There are two types of stereoisomerism, **optical isomerism** and **geometrical isomerism** (***cis-trans*** **isomerism**). It is not easy to define them, but their meanings will become clear as the study of stereochemistry progresses. Even so, it is highly desirable to have some idea about their meanings at this stage, and so the following summaries are given.

Optical isomerism is characterised by compounds having the same structure but different configurations, and because of their *molecular asymmetry* these compounds rotate the plane of polarisation of plane-polarised light. Optical isomers have similar physical and chemical properties; the most marked difference between them is their action on plane-polarised light (see §12. I). Optical isomers may rotate the plane of polarisation by *equal* and *opposite* amounts; these optical isomers are **enantiomorphs** (see §2). On the other hand, some optical isomers may rotate the plane of polarisation by *different* amounts; these are **diastereoisomers** (see §7b). Finally, some optical isomers may possess no rotation at all; these are diastereoisomers of the **meso-type** (see §7d).

Geometrical isomerism is characterised by compounds having the same structure but different configurations, and because of their *molecular symmetry* these compounds do *not* rotate the plane of polarisation of plane-polarised light. Geometrical isomers differ n all their physical and in many of their chemical properties. They an also exhibit optical isomerism if the structure of the molecule,

apart from giving rise to geometrical isomerism, is also asymmetric. In general, geometrical isomerism involves molecules which can assume different stable configurations, the ability to do so being due, *e.g.*, to the presence of a double bond, a ring structure, or the spatial effect (see Ch. IV and V).

§2. Optical isomerism. It has been found that only those structures, crystalline or molecular, which are *not* superimposable on their mirror images, are optically active. Such structures may be *dissymmetric*, or *asymmetric*. Asymmetric structures have no elements of symmetry at all, but dissymmetric structures, although possessing some elements of symmetry, are nevertheless still capable of existing in two forms (one the mirror image of the other) which are not superimposable. To avoid unnecessary complications, we shall use the term asymmetric to cover both cases (of asymmetry and dissymmetry).

Optical activity due to crystalline structure. There are many substances which are optically active in the solid state only, *e.g.*, quartz, sodium chlorate, benzil, etc. Let us consider quartz, the first substance shown to be optically active (Arago, 1811). Quartz exists in two crystalline forms, one of which is dextrorotatory and the other lævorotatory. These two forms are mirror images and are not superimposable. Such pairs of crystals are said to be *enantiomorphous* (quartz crystals are actually hemihedral and are mirror images). X-ray analysis has shown that the quartz crystal lattice is built up of silicon and oxygen atoms arranged in left- and right-handed spirals. One is the mirror image of the other, and the two are not superimposable. When quartz crystals are fused, the optical activity is lost. Therefore the optical activity is entirely due to the *asymmetry of the crystalline structure*, since fusion brings about only a physical change. Thus we have a gr[illegible] active only so long as they [illegible] or solution in a solvent cause[illegible]

Optical activity due to molecular structure. There are many compounds which are optically active in the solid, fused, gaseous, or dissolved state, *e.g.*, glucose, tartaric acid, etc. In this case the optical activity is entirely due to the *asymmetry of the molecular structure* (see, however, §11). The original molecule and its non-superimposable mirror image are known as *enantiomorphs* (this name is taken from crystallography) or *optical antipodes*. They are also often referred to as *optical isomers*, but there is a tendency to reserve this term to denote *all* isomers which have the same structural formula but different configurations (see §1).

Properties of enantiomorphs. It appears that enantiomorphs are identical physically except in two respects:

(i) their manner of rotating polarised light; the rotations are equal but opposite.

(ii) the absorption coefficients for dextro- and lævocircularly polarised light are different; this difference is known as *circular dichroism* or the *Cotton effect* (see also §8. III).

The crystal forms of enantiomorphs may be mirror images of each other, *i.e.*, the crystals themselves may be enantiomorphous, but this is unusual [see also §10 (i)]. Enantiomorphs are similar chemically, but their rates of reaction with other optically active compounds are usually different [see §10 (vii)]. They may also be different physiologically, *e.g.*, (+)-histidine is sweet, (−)- tasteless; (−)-nicotine is more poisonous than (+)-.

§3. **The tetrahedral carbon atom.** In 1874, van't Hoff and Le Bel, independently, gave the solution to the problem of optical isomerism in organic compounds. van't Hoff proposed the theory that if the four valencies of the carbon atom are arranged tetrahedrally (not necessarily regular) with the carbon atom at the centre, then all the cases of isomerism known are accounted for. Le Bel's theory is substantially the same as van't Hoff's, but differs in that whereas van't Hoff believed that the valency distribution was definitely tetrahedral and fixed as such, Le Bel believed that the valency directions were not rigidly fixed, and did not specify the tetrahedral arrangement, but thought that *whatever* the spatial arrangement, the molecule *Cabde* would be *asymmetric*. Later work has shown that van't Hoff's theory is more in keeping with the facts (see below). Both van't Hoff's and Le Bel's theories were based on the assumption that the four hydrogen atoms in methane are equivalent; this assumption has been shown to be correct by means of chemical and physico-chemical methods. Before the tetrahedral was proposed, it was believed that the four carbon valencies were planar, with the carbon atom at the centre of a square (Kekulé, 1858).

Pasteur (1848) stated that all substances fell into two groups, those which were superimposable on their mirror images, and those which were not. In substances such as quartz, optical activity is due to the dissymmetry of the *crystal* structure, but in compounds like sucrose the optical activity is due to *molecular* dissymmetry. Since it is impossible to have molecular dissymmetry if the molecule is flat, Pasteur's work is based on the idea that molecules are three-dimensional and arranged dissymmetrically. A further interesting point in this connection is that Pasteur quoted an irregular

tetrahedron as one example of a dissymmetric structure. Also, Paterno (1869) had proposed *tetrahedral models* for the structure of the isomeric compounds $C_2H_4Cl_2$ (at that time it was thought that there were three isomers with this formula: one ethylidene chloride and two ethylene chlorides).

§3a. Evidence for the tetrahedral carbon atom. The molecule CX_4 constitutes a five-point system, and since the four valencies of carbon are equivalent, their disposition in space may be assumed to be symmetrical. Thus there are three symmetrical arrangements possible for the molecule CX_4, one planar and two solid—pyramidal and tetrahedral. By comparing the number of isomers that have been prepared for a given compound with the number predicted by the above three spatial arrangements, it is possible to decide which one is correct.

Compounds of the types Ca_2b_2 *and* Ca_2bd. Both of these are similar, and so we shall only discuss molecule Ca_2b_2.

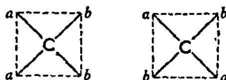


FIG. 2.1.

(i) If the molecule is planar, then *two* forms are possible (Fig. 1). This planar configuration can be either square or rectangular; in each case there are two forms only.

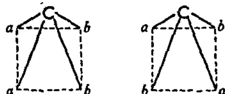


FIG. 2.2.

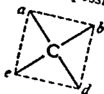
(ii) If the molecule is pyramidal, then *two* forms are possible (Fig. 2). There are only two forms, whether the base is square or rectangular.

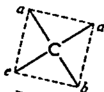


(iii) If the molecule is tetrahedral, then only *one* form is possible (Fig. 3; the carbon atom is at the centre of the tetrahedron).

In practice, only one form is known for each of the compounds of the types Ca_2b_2 and Ca_2bd; this agrees with the tetrahedral configuration.

Compounds of the type Cabde. (i) If the molecule is planar, then *three* forms are possible (Fig. 4).





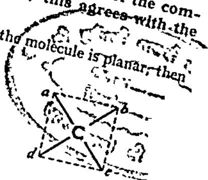
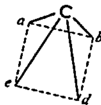


Fig 2.4.

(ii) If the molecule is pyramidal, then *six* forms are possible; there are three pairs of enantiomorphs. Each of the forms in Fig. 4, drawn as a pyramid, is not superimposable on its mirror image, *e.g.*, Fig. 5 shows one pair of enantiomorphs.



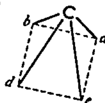
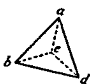


Fig. 2.5.

(iii) If the molecule is tetrahedral, there are *two* forms possible, one related to the other as object and mirror image, which are not superimposable, *i.e.*, the tetrahedral configuration gives rise to one pair of enantiomorphs (Fig. 6).



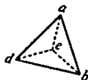
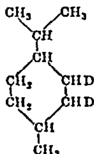


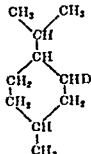
Fig. 2.6.

In practice, compounds of the type *Cabde* give rise to only one pair of enantiomorphs; this agrees with the tetrahedral configuration.

When a compound contains four different groups attached to a carbon atom, that carbon atom is said to be asymmetric (actually, of course, it is the *group* which is asymmetric; a *carbon atom* cannot be asymmetric). The majority of optically active compounds (organic) contain one or more asymmetric carbon atoms. It should be remembered, however, that the *essential* requirement for optical activity is the *asymmetry of the molecule.* A molecule may contain two or more asymmetric carbon atoms and still not be optically active (see, *e.g.*, §7d).

Isotopic asymmetry. In the optically active compound C*abde*, the groups *a*, *b*, *d* and *e* (which may or may not contain carbon) are all different, but two or more may be *structural* isomers, *e.g.*, propyl*iso*propylmethanol is optically active. The substitution of hydrogen by deuterium has also been investigated in recent years to ascertain whether these two atoms are sufficiently different to give rise to optical isomerism. The earlier work gave conflicting results, *e.g.*, Clemo *et al.* (1936) claimed to have obtained a small rotation for α-pentadeuterophenylbenzylamine, $C_6D_5{\cdot}CH(C_6H_5){\cdot}NH_2$, but this was disproved by Adams *et al.* (1938). Erlenmeyer *et al.* (1936) failed to resolve $C_6H_5{\cdot}CH(C_6D_5){\cdot}CO_2H$, and Ives *et al.* (1948) also failed to resolve a number of deutero-compounds, one of which was $C_6H_5{\cdot}CH_2{\cdot}CHD{\cdot}CO_2H$. More recent work, however, appears to be conclusive in favour of optical activity, *e.g.*, Eliel (1949) prepared optically active methylphenyldeuteromethane, $CH_3{\cdot}CHD{\cdot}C_6H_5$, by reducing optically active methylphenylmethyl chloride, $CH_3{\cdot}CHCl{\cdot}C_6H_5$, with lithium aluminium deuteride; Ross *et al.* (1956) have prepared (—)-2-deuterobutane by reduction of (—)-2-chlorobutane with lithium aluminium deuteride; and Alexander *et al.* (1949) reduced *trans*-2-*p*-menthene with deuterium (Raney nickel catalyst) and obtained a 2:3-dideutero-*trans*-*p*-menthane (I) that was slightly lævorotatory. Alexander (1950) also reduced (—)-menthyl toluene-*p*-sulphonate and obtained an optically active 3-deutero-*trans*-*p*-menthane (II).





whereby the carboxyl and the carbonamide groups were interchanged, the product was (—)-*iso*propylmalonamic acid. It is most important to note that in this series of reactions no bond connected to the asymmetric carbon atom was ever broken (for an explanation, see Walden Inversion, Ch. III).

This change from one enantiomorph into the other is in agreement with the tetrahedral theory. At the same time, this series of reactions shows that optical isomers have identical structures, and so the difference must be due to the spatial arrangement.

(iii) X-ray crystallography, dipole moment measurements, absorption spectra and electron diffraction studies show that the four valencies of carbon are arranged tetrahedrally with the carbon atom inside the tetrahedron.

It should be noted in passing that the tetrahedra are not regular unless four identical groups are attached to the central carbon atom; only in this case are the four bond lengths equal. In all other cases the bond lengths will be different, the actual values depending on the nature of the atoms joined to the carbon atom (see §15b I).

§4. Two postulates underlie the tetrahedral theory.

(i) **The principle of constancy of the valency angle.** Mathematical calculation of the angle subtended by each side of a regular tetrahedron at the central carbon atom (Fig. 9) gives a value of 109° 28′. Originally, it was postulated (van't Hoff) that the valency angle was fixed at this value. It is now known, however, that the valency angle may deviate from this value. The four valencies of carbon are formed by hybridisation of the $2s^2$ and $2p^2$ orbitals, *i.e.*, there are four sp^3 bonds (see Vol. I, Ch. II). Quantum mechanical calculations show that the four carbon valencies in the molecule Ca_4 are equivalent and directed towards the four corners of a regular tetrahedron. Furthermore, quantum mechanical calculations require the carbon bond angles to be close to the tetrahedral value, since change from this value is associated with loss in bond strength and consequently decrease in stability. According to Coulson *et al.* (1949), calculation has shown that the *smallest* valency angle that one can reasonably expect to find is 104°. It is this value which is found in the *cyclo*propane and *cyclo*butane rings, these molecules being relatively unstable because of the "bent" bonds (Coulson; see Baeyer Strain Theory, Vol. I, Ch. XIX).

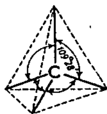


FIG 2.9.

(ii) **The principle of free rotation about a single bond.** Originally, it was believed that internal rotation about a single bond was completely free. When the thermodynamic properties

were first calculated for ethane on the assumption that there was complete free rotation about the carbon–carbon single bond, the results obtained were in poor agreement with those obtained experimentally. This led Pitzer *et al.* (1936) to suggest that there was

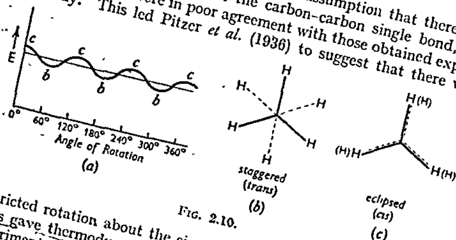


Fig. 2.10.

restricted rotation about the single bond, and calculations on this basis gave thermodynamic properties in good agreement with the experimental ones. The potential energy curve obtained for ethane, in which one methyl group is imagined to rotate about the C—C bond as axis with the other group at rest, is shown in Fig. 10 (*a*). Had there been complete free rotation, the graph

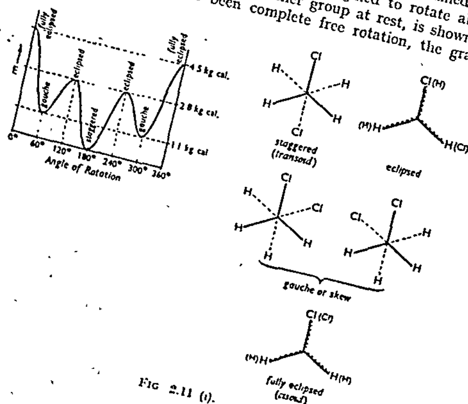


Fig 2.11 (i).

would have been a horizontal straight line. Fig. 10 (*b*) [this *projection formula* is obtained by viewing the molecule along the bonding line of the two carbon atoms] represents the *trans* or **staggered** form in which the hydrogen atoms (on the two carbon atoms) are as far apart as possible. Fig. 10 (*c*) represents the *cis* or **eclipsed** form in which the hydrogen atoms are as close together as possible. It can be seen from the graph that the eclipsed form has a higher potential energy than the staggered, and the actual difference has been found to be (by calculation) about 2·85 kg.cal./mole. The value of this potential energy barrier is too low to permit the isolation of each form by chemical methods.

Now let us consider the case of ethylene chloride. According to Bernstein (1949), the potential energy of ethylene chloride undergoes the changes shown in Fig. 2.11 (i) when one CH_2Cl group is rotated about the C—C bond with the other CH_2Cl at rest. There are two positions of minimum energy, one corresponding to the staggered (transoid) form and the other to the gauche

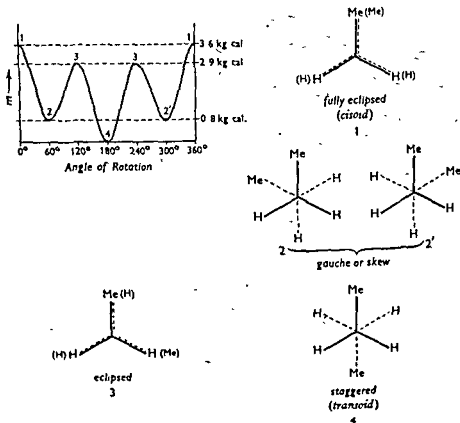


Fig. 2.11 (ii).

(skew) form, the latter possessing approximately 1·1 kg.cal. more than the former. The fully eclipsed (cisoid) form possesses about 4·5 kg.cal. more energy than the staggered form and thus the latter is the preferred form, *i.e.*, the molecule is largely in this form. Dipole moment studies show that this is so in practice, and also show (as do Raman spectra studies) that the ratio of the two forms varies with the temperature. Furthermore, infra-red, Raman spectra and electron diffraction studies have shown that the gauche form is also present. According to Mizushima *et al.* (1938), only the staggered form is present at low temperatures.

The problem of internal rotation about the central C—C bond in *n*-butane is interesting, since the values of the potential energies of the various forms have been used in the study of cyclic compounds (see *cyclo*hexane, §11. IV). The various forms are shown in Fig. 2.11 (ii), and if the energy content of the staggered form is taken as zero, then the other forms have the energy contents shown (Pitzer, 1951).

From the foregoing account it can be seen that, in theory, there is no free rotation about a single bond. In practice, however, it may occur if the potential barriers of the various forms do not differ by more than about 10 kg.cal./mole. Free rotation about a single bond is generally accepted in *simple* molecules. Restricted rotation, however, may occur when the molecule contains groups large enough to impede free rotation, *e.g.*, in *ortho*-substituted diphenyls (see Ch. V). In some cases resonance can give rise to restricted rotation about a "single" bond.

§4a. Conformational analysis. Molecules which can form isomers by rotation about single bonds are called **flexible molecules**, and the different forms taken up are known as different **conformations**. The terms *rotational isomers* and *constellations* have also been used in the same sense as conformations.

Various definitions have been given to the term *conformation* (which was originally introduced by W. N. Haworth, 1929). In its widest sense, conformation has been used to describe different spatial arrangements of a molecule which are not superimposable. This means, in effect, that the terms *conformation* and *configuration* are equivalent. There is, however, an important difference in meaning between these terms. The definition of configuration, in the classical sense (§1), does not include the problem of the internal forces acting on the molecule. The term conformation, however, is the spatial arrangement of the molecule when all the internal forces acting on the molecule are taken into account. In this more restricted sense, the term conformation is used to designate

different spatial arrangements arising by twisting or rotation of bonds of a *given* configuration (used in the classical sense).

The existence of potential energy barriers between the various conformations shows that there are internal forces acting on the molecule. The nature of these interactions that prevent free rotation about single bonds, however, is not completely clear. According to one theory, the hindering of internal rotation is due to dipole–dipole forces. Calculation of the dipole moment of ethylene chloride on the assumption of free rotation gave a value not in agreement with the experimental value. Thus free rotation cannot be assumed, but on the assumption that there is interaction between the two groups through dipole–dipole attractive or repulsive forces, there will be preferred conformations, *i.e.*, the internal rotation is not completely free. This restricted rotation is shown by the fact that the dipole moment of ethylene chloride increases with temperature; in the staggered form the dipole moment is zero, but as energy is absorbed by the molecule, rotation occurs to produce finally the eclipsed form in which the dipole moment is a maximum. Further work, however, has shown that factors other [illegible]
in opposing [illegible]
i.e., repulsi[illegible]
groups) when they are brought into close proximity (cf. the van der Waals forces, §2. I). The existence of steric repulsion may be illustrated by the fact that although the bond moment of C—Cl is greater than that of C—Br, the energy difference between the eclipsed and staggered conformations of ethylene chloride is less than that of ethylene bromide. Furthermore, if steric repulsion does affect internal rotation, then in the ethylene halides, steric repulsion between the hydrogen and halogen atoms, if sufficiently large, will give rise to two other potential energy minima (these correspond to the two gauche forms, and these have been shown to be present; see Fig. 2.11 (i), §4).

Other factors also affect stability of the various conformations. Staggered and gauche forms always exist in molecules of the type $CH_2Y{\cdot}CH_2Z$ (where Y and Z are Cl, Br, I, CH_3, etc.), and usually the staggered form is more stable than the gauche. In a molecule such as ethylene chlorohydrin, however, it is the gauche form which is more stable than the staggered, and this is due to the fact that intramolecular hydrogen bonding is possible in the former but not in the latter.

Still other causes have been proposed to account for the absence of complete free rotation about a single bond. *E.g.*, since the energy required to rotate one methyl group about the C—C single

bond in ethane is greater than expected from steric repulsion between the hydrogen atoms, it has been suggested that the cause is *second-order hyperconjugation* (see Vol. I):

$$H_3{\equiv}C{-}C{\equiv}H_3 \longleftrightarrow H_3{=}C{=}C{=}H_3$$

Thus the C—C bond acquires a small amount of double-bond character and consequently offers more resistance to rotation than had it been a pure single bond. Calculations, however, have shown that the resistance to the rotation is too small to be accounted for by hyperconjugation.

When the stability of a molecule is decreased by internal forces produced by interaction between constituent parts, that molecule is said to be under **steric strain**. There are three sources of steric strain, *i.e.*, the internal forces may arise from three different causes, *viz.*, (i) repulsion between non-bonded atoms, (ii) dipole interactions, and (iii) distortion of bond-angles. Which of these plays the predominant part depends on the nature of the molecule in question. This study of the existence of preferred conformations in molecules, and the relating of physical and chemical properties of a molecule to its preferred conformation, is known as **conformational analysis.** The energy differences between the various conformations determine which one is the most stable, and the ease of transformation depends on the potential energy barriers that exist between these conformations. It should be noted that the molecule, in its *unexcited* state, will exist largely in the conformation of lowest energy content. If, however, the energy differences between the various conformations are small, then when *excited*, the molecule can take up a less favoured conformation, *e.g.*, during the course of reaction with other molecules (see §11. IV).

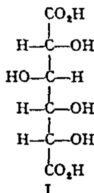
Because of the different environments, a reactive centre may have in different conformations, conformation will therefore affect the course and rate of reactions involving this centre (see §11. IV).

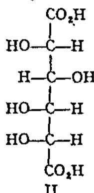
Many methods are now used to investigate the conformation of molecules, *e.g.*, thermodynamic calculations, dipole moments, electron and X-ray diffraction, infra-red and Raman spectra, and chemical methods.

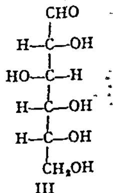
§5. Conventions used in stereochemistry. The original method of indicating enantiomorphs was to prefix each one by *d* or *l* according as it was dextrorotatory or lævorotatory. van't Hoff (1874) introduced a + and − notation for designating the configuration of an asymmetric carbon atom. He used mechanical models (built of tetrahedra), and the + and − signs were given by observing the tetrahedra of the mechanical model from the centre

of the model. Thus a molecule of the type *Cabd·Cabd* may be designated + +, − −, and + −. E. Fischer (1891) pointed out that this + and − notation can lead to wrong interpretations when applied to molecules containing more than two asymmetric carbon atoms (the signs given to each asymmetric carbon atom depend on the point of observation in the molecule). Fischer therefore proposed the use of plane projection diagrams of the mechanical models instead of the + and − system.

Fischer, working on the configurations of the sugars (see §1. VII), obtained the plane formulæ I and II for the enantiomorphs of saccharic acid, and **arbitrarily** chose I for dextrorotatory saccharic

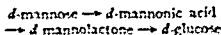






acid, and called it *d*-saccharic acid. He then, from this, deduced formula III for *d*-glucose. Furthermore, Fischer thought it was more important to indicate stereochemical relationships than merely to indicate the actual direction of rotation. He therefore proposed that the **prefixes *d* and *l* should refer to stereochemical relationships and not to the direction of rotation of the compound.** For this scheme to be self-consistent (among the sugars) it is necessary to choose *one* sugar as standard and then refer all the others to it. Fischer apparently intended to use the scheme whereby the compounds derived from *a given aldehyde sugar* should be designated according to the *direction of rotation of the parent aldose.*

Natural mannose is dextrorotatory. Hence natural mannose will be *d*-mannose, and all derivatives of *d*-mannose, *e.g.*, mannonic acid, mannitol, mannose phenylhydrazone, etc., will thus belong to the *d*-series. Natural glucose is dextrorotatory. Hence natural glucose will [illegible] belong to the *d*-series. F[illegible] natural mannose into natural glucose as follows:



Since natural glucose is *d*-glucose (according to Fischer's scheme), the prefix *d* for natural glucose *happens* to agree with its dextrorotation (with *d*-mannose as standard). Natural fructose can also be prepared from natural mannose (or natural glucose), and so will be *d*-fructose. Natural fructose, however, is lævorotatory, and so is written as *d*(—)-fructose, the symbol *d* indicating its *stereochemical* relationship to the parent aldose glucose, and the symbol — placed in parentheses before the name indicating the *actual direction of rotation*.

More recently the symbols *d* and *l* have been replaced by D and L for configurational relationships, *e.g.*, L(+)-lactic acid. Also, when dealing with compounds that cannot be referred to an arbitrarily chosen standard, (+)- and (—)- are used to indicate the sign of the rotation. The prefixes *dextro* and *lævo* (without hyphens) are also used.

Fischer's proposal to use *each aldose* as the arbitrary standard for its derivatives leads to some difficulties, *e.g.*, natural arabinose is dextrorotatory, and so is to be designated D-arabinose. Now natural arabinose (D-arabinose) can be converted into mannonic acid which, if D-arabinose is taken as the parent aldose, will therefore be D-mannonic acid. This same acid, however, can also be obtained from L-mannose, and so should be designated as L-mannonic acid. Thus in cases such as this the use of the symbol D or L will depend on the *historical order* in which the stereochemical relationships were established. This, obviously, is an unsatisfactory position, which was realised by Rosanoff (1906), who showed that if the enantiomorphs of glyceraldehyde (a molecule which contains only *one* asymmetric carbon atom) are chosen as the (arbitrary) standard, then a satisfactory system for correlating stereochemical relationships can be developed. He also proposed that the formula of dextrorotatory glyceraldehyde should be written as in Fig. 12 (*c*), in order that the arrangement of its asymmetric carbon atom should agree with the arrangement of C_5 in Fischer's projection formula for natural glucose (see formula III above).

It is of great interest to note in this connection that in 1906 the active forms of glyceraldehyde had not been isolated, but in 1914 Wohl and Momber separated DL-glyceraldehyde into its enantiomorphs, and in 1917 they showed that dextrorotatory glyceraldehyde was stereochemically related to natural glucose, *i.e.*, with D(+)-glyceraldehyde as arbitrary standard, natural glucose is D(+)-glucose.

The accepted convention for drawing D(+)-glyceraldehyde—the agreed (*arbitrary*) standard—is shown in Fig. 12 (*a*). The tetrahedron is drawn so that three corners are imagined to be *above*

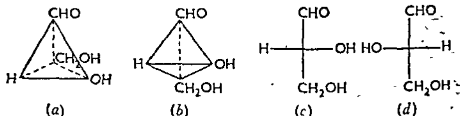
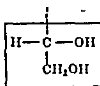
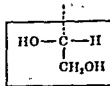


Fig. 2.12

the plane of the paper, and the fourth *below* the plane of the paper. Furthermore, the spatial arrangement of the four groups joined to the central carbon atom *must be placed as shown in Fig.* 12 (*a*), *i.e.*, **the accepted convention for drawing D(+)-glyceraldehyde places the hydrogen atom at the left and the hydroxyl group at the right, with the aldehyde group at the top corner.** Now imagine the tetrahedron to rotate about the horizontal line joining H and OH until it takes up the position shown in Fig. 12 (*b*). This is the *conventional* position for a tetrahedron, groups joined to *full horizontal* lines being *above* the plane of the paper, and those joined to *broken vertical* lines being *below* the plane of the paper. The *conventional plane-diagram* is obtained by drawing the full horizontal and broken vertical lines of Fig. 12 (*b*) as full lines, placing the groups as they appear in Fig 12 (*b*), and taking the asymmetric carbon atom to be at the point where the lines cross. Although Fig. 12 (*c*) is a plane-diagram, it is most important to remember that horizontal lines represent groups above the plane, and vertical lines groups below the plane of the paper. Fig. 12 (*d*) represents the plane-diagram formula of L(−)-glyceraldehyde; here *the hydrogen atom is to the right and the hydroxyl group to the left.* Thus any compound that can be prepared from, or converted into, D(+)-glyceraldehyde will belong to the D-series. Similarly, any compound that can be prepared from, or converted into, L(−)-glyceraldehyde will belong to the L-series. When representing relative configurational relationship of molecules containing more than one asymmetric carbon atom, *the asymmetric carbon atom of glyceraldehyde is always drawn at the bottom*, the rest of the molecule being built up from this unit.



D-series



L-series

Thus we have a scheme of classification of *relative* configurations based on D(+)-glyceraldehyde as *arbitrary* standard. Even on this basis confusion is still possible in relating configurations to the standard (see later). To overcome these difficulties, Cahn and Ingold (1951) have suggested a "sequence rule". These authors propose that the groups round each asymmetric carbon atom should be placed in order of decreasing atomic numbers of certain selected Sets of their atoms, and correlated with the OH, CHO, CH_2OH, and H of glyceraldehyde. A Set comprises certain atoms which are specified by a rule. If the order of the groups is clockwise when viewed from the side opposite to the *hydrogen* atom (or the group correlated with this hydrogen atom), the configuration is D, and if

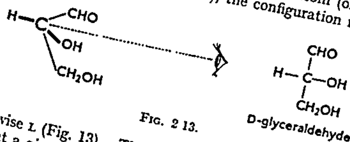
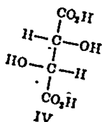
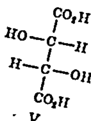


FIG. 2 13.

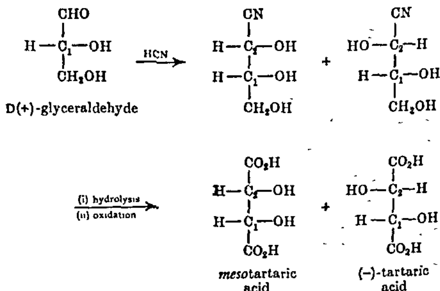
anticlockwise L (Fig. 13). The symbols D and L refer to configurations about a *given* atom which must be numbered or lettered, *e.g.*, natural dextrorotatory tartaric acid will be αD:α'D-tartaric acid (see appropriate reading reference for further details).

Until recently there was no way of determining, with certainty, the *absolute* configuration of molecules. *Arbitrary choice* makes the configuration of D(+)-glyceraldehyde have the hydrogen to the left and the hydroxyl to the right. Bijvoet *et al.* (1951), however, have shown by X-ray analysis of sodium rubidium tartrate that it is possible to differentiate between the two optically active forms, *i.e.*, it is possible to determine the *absolute* configuration of these two enantiomorphs. These authors showed that natural dextrorotatory tartaric acid has the configuration assigned to it by Fischer (who correlated its configuration with that of the saccharic acids). The configurations of the tartaric acids are a troublesome problem.

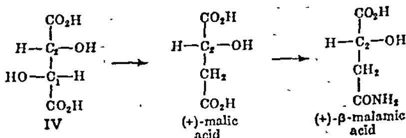


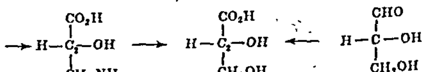


Fischer wrote the configuration of natural dextrorotatory tartaric acid as IV. If we use the convention of writing the glyceraldehyde unit at the bottom, then IV is L(+)-tartaric acid and V D(−)-tartaric acid. This relationship (to glyceraldehyde) is confirmed



by the conversion of D(+)-glyceraldehyde into lævorotatory tartaric acid *via* the Kiliani reaction (see Vol. I). Thus (−)-tartaric acid is D(−)-tartaric acid (V). On the other hand, (+)-tartaric acid can be converted into D(−)-glyceric acid, and so (+)-tartaric acid is D(+)-tartaric acid (IV). The confusion arises because in one set



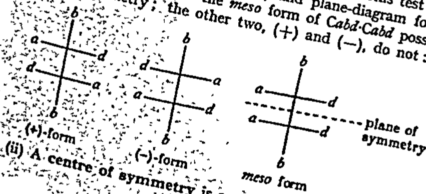


of reactions the configuration of C_1 has been correlated with D(+)-glyceraldehyde, whereas in the other set of reactions it is the configuration of C_2 that has been correlated. The scheme proposed by Cahn and Ingold clears up the difficulty by relating *each individual asymmetric carbon atom* to D(+)-glyceraldehyde according to the sequence rule. On this basis, natural dextrorotatory tartaric acid is thus (+)-αD:α'D-tartaric acid (IV), D(+)-tartaric acid.

Thus an important problem when dealing with optically active compounds is to correlate configurations with a standard (whether arbitrary or absolute). The configurations of many classes of natural products have now been correlated with glyceraldehyde (see carbohydrates, §1. VII; terpenes, §23e. VIII; steroids, §4b. XI; amino-acids, §4. XIII). Various methods have been used for correlation of configuration: chemical methods (*e.g.*, as described for tartaric acid), the study of optical rotations (*e.g.*, steroids), and the study of quasi-racemic compounds (see §9a).

§6. Elements of symmetry. The test of superimposing a formula (tetrahedral) on its mirror image definitely indicates whether the molecule is symmetrical or not; it is asymmetric if the two forms are not superimposable. The most satisfactory way in which superimposability may be ascertained is to build up models of the molecule and its mirror image. Usually this is not convenient, and so, in practice, one determines whether the molecule possesses (i) a plane of symmetry, (ii) a centre of symmetry, or (iii) an alternating axis of symmetry. If the molecule contains at least one of these elements of symmetry, the molecule is symmetrical; if none of these elements of symmetry is present, the molecule is asymmetric.

(i) A **plane of symmetry** divides a molecule in such a way that points (atoms or groups of atoms) on the one side of the plane form mirror images of those on the other side. This test may be applied to both solid (tetrahedral) and plane-diagram formulae, *e.g.*, the plane-formula of the *meso* form of *Cabd·Cabd* possesses a plane of symmetry; the other two, (+) and (−), do not:



(ii) A **centre of symmetry** is a point from which lines, when

drawn on one side and produced an equal distance on the other side, will meet exactly similar points in the molecule. This test can be satisfactorily applied only to three-dimensional formulae, particularly those of ring systems, *e g.*, 2:4-dimethyl*cyclo*butane-1.3-dicarboxylic acid *(Fig. 14). The form shown possesses a centre of* symmetry which is the centre of the ring. This form is therefore optically inactive.

Another example we shall consider here is that of dimethyldiketopiperazine; this molecule can exist in two geometrical isomeric

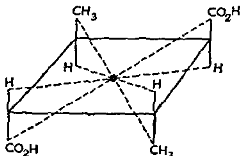
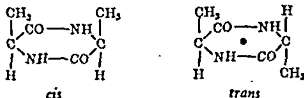


Fig. 2.14.

forms, *cis* and *trans* (see also §11. IV). The *cis* isomer has no elements of symmetry and can therefore exist in two enantiomorphous forms; both are known. The *trans* isomer has a centre of symmetry and is therefore optically inactive.



It is important to note that only *even-membered* rings can possibly possess a centre of symmetry.

(iii) **Alternating axis of symmetry.** A molecule possesses an *n*-fold alternating axis of symmetry if, when rotated through an angle of $360°/n$ about this axis and then followed by reflection in a plane perpendicular to the axis, the molecule is the same as it was in the starting position. Let us consider the molecule shown in *Fig. 15 (a)* [1:2:3:4-tetramethyl*cyclo*butane]. This contains a four-fold alternating axis of symmetry. Rotation of *(a)* through 90° about axis AB which passes through the centre of the ring perpendicular to its plane gives *(b)*, and reflection of *(b)* in the

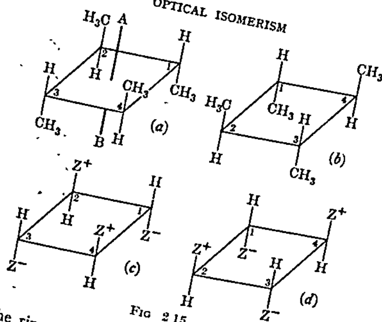
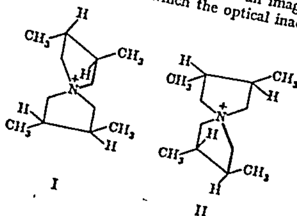


Fig 2 15.

plane of the ring gives (*a*). It also happens that this molecule possesses two vertical planes of symmetry (through each diagonal of the ring), but if the methyl groups are replaced alternately by the asymmetric groups (+)—$CH(CH_3)\cdot C_2H_5$ and (−)—$CH(CH_3)\cdot C_2H_5$, represented by Z+ and Z− respectively, the resulting molecule (Fig. 15 *c*) now has no planes of symmetry. Nevertheless, this molecule is *not* optically active since it does possess a four-fold alternating axis of symmetry [reflection of (*d*) (which is produced by rotation of (*c*) through 90° about the vertical axis) in the plane of the ring gives (*c*); it should be remembered that the reflection of a (+)-form is the (−)-form].

The *cyclo*butane derivative (*c*) given above to illustrate the meaning of an alternating axis of symmetry is an imaginary molecule. No compound was known in which the optical inactivity was due



to the existence of *only* an alternating axis until McCasland and Proskow (1956) prepared such a molecule for the first time. This is a spiro-type of molecule (§7. V), *viz.*, 3:4:3′:4′-tetramethylspiro-*(1:1′)-dipyrrolidinium p-toluenesulphonate, I (the p-toluenesulphonate ion has been omitted).* This molecule is discussed in some detail in §2a. VI, but here we shall examine it for its alternating axis of symmetry. Molecule I is superimposable on its mirror image and hence is not optically active. It does not contain a plane or centre of symmetry, but it does contain a four-fold alternating axis of symmetry. To show the presence of this axis, if I rotated through 90° about the co-axis of both rings, II is obtained. Reflection of II through the central plane (*i.e.*, through the N atom) perpendicular to this axis gives a molecule identical and coincident with I.

In practice one decides whether a molecule is symmetrical or not by looking only for a plane or centre of symmetry, since no *natural* compound has yet been found to have an alternating axis of symmetry. The presence of two or more asymmetric carbon atoms will definitely give rise to optical isomerism, but nevertheless *some* isomers may not be optically active because these *molecules as a whole* are not asymmetric (see §7d).

§7. The number of isomers in optically active compounds. *The number of optical isomers that can theoretically* be derived from a molecule containing one or more asymmetric carbon atoms is of fundamental importance in stereochemistry.

§7a. Compounds containing one asymmetric carbon atom. With the molecule *Cabde* only two optical isomers are possible, and these are related as object and mirror image, *i.e.*, there is one pair of enantiomorphs, *e.g.*, D- and L-lactic acid. If we examine an *equimolecular* mixture of dextrorotatory and lævorotatory lactic acids, we shall find that the mixture is optically inactive. This is to be expected, since enantiomorphs have equal but opposite rotatory power. Such a mixture (of equimolecular amounts) is said to be **optically inactive by external compensation**, and is known as a **racemic modification** (see also §9). A compound which is optically inactive by external compensation is known as the **racemic compound** and is designated as *r*-, (±)- or DL-, *e.g.*, *r*-tartaric acid, (±)-limonene, DL-lactic acid.

Thus a compound containing *one* asymmetric carbon atom can exist in *three* forms: (+)-, (−), and (±).

Conversion of molecule Ca_2bd *into Cabde.* Let us consider as an example the bromination of propionic acid to give α-bromopropionic acid.

$$CH_3 \cdot CH_2 \cdot CO_2H \xrightarrow{Br_2/P} CH_3 \cdot CHBr \cdot CO_2H$$

II and III (Fig. 16) are enantiomorphs, and since molecule [illegible] symmetrical about its vertical axis, it can be anticipated from the

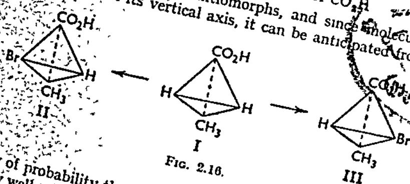


FIG. 2.16.

theory of probability that either hydrogen atom should be replaced equally well to give (±)-α-bromopropionic acid. This actually does occur in practice.

§7b. Compounds containing two different asymmetric carbon atoms. When we examine the molecule *Cabd·Cabe*, *e.g.*, α : β-dibro[illegible]ic acid, $CH_3 \cdot CHBr \cdot CHBr \cdot CO_2H$, we find that there [illegible]patial arrangements for this type of molecule [illegible] enantiomorphs (the configurations of *both*

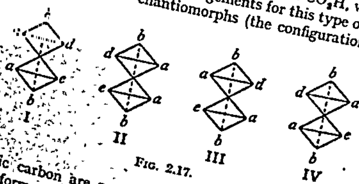


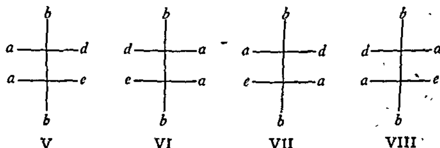
FIG. 2.17.

asymmetric carbon are reversed), and an equimolecular mixture of them forms a racemic modification; similarly for III and IV. Thus there are six forms in all for a compound of the type *Cabd·Cabe*: two pairs of enantiomorphs and two racemic modifications.

I and III are not identical in configuration and are not mirror images (the configuration of *one* of the two asymmetric carbon atoms is reversed); they are known as **diastereoisomers**, *i.e.*, they are optical isomers but not enantiomorphs (mirror images). Diastereoisomers differ in physical properties such as melting point, density, solubility, dielectric constant and specific rotation. Chemically they are similar, but their rates of reaction with other

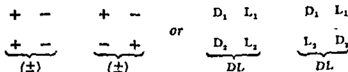
optically active compounds are different (*cf.* the properties of enantiomorphs, §2).

The plane-diagrams of molecules I–IV (Fig. 17) will be V–VIII, respectively, as shown. It should be remembered that groups

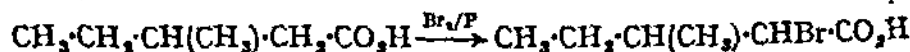


joined to horizontal lines lie above the plane of the paper, and those joined to vertical lines lie below the plane of the paper (§5).

Instead of writing down all the possible configurations, the number of optical isomers for a compound of the type *Cabd·Cabe* may be obtained by indicating the *configuration* of each asymmetric carbon atom by the symbol + or −, or by D or L; thus:

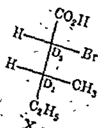


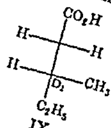
Conversion of molecule Ca_2b·*Cabe into Cabd·Cabe.* Let us consider the bromination of β-methylvaleric acid to give α-bromo-β-methylvaleric acid.

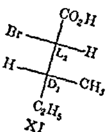


$$CH_3{\cdot}CH_2{\cdot}CH(CH_3){\cdot}CH_2{\cdot}CO_2H \xrightarrow{Br_2/P} CH_3{\cdot}CH_2{\cdot}CH(CH_3){\cdot}CHBr{\cdot}CO_2H$$

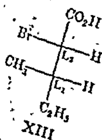
β-Methylvaleric acid contains *one* asymmetric carbon atom, but the bromine derivative contains *two*. Let us first consider the case where the configuration of the asymmetric carbon atom in the starting material is D_1 (IX). Bromination of this will produce molecules X and XI; these are diastereoisomers and are produced in *unequal* amounts. This is to be anticipated; the two α-hydrogen atoms are not symmetrically placed with respect to the lower half of the molecule, and consequently different rates of substitution can be expected. In the same way, bromination of the starting material in which the configuration of the asymmetric carbon atom is L_1 (XII) leads to the formation of a mixture of diastereoisomers (XIII and XIV) in unequal amounts. One can expect, however, that the amount of XIII produced from XII would be the same as

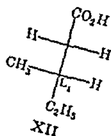


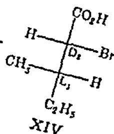




that of X from IX since, in both cases, the positions of the bromine atoms with respect to the methyl group are the same. Similarly, the amount of XIV from XII will be the same as that of XI from

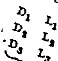


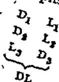


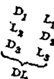


IX. Thus bromination of (±)-β-methylvaleric acid will result in a mixture of four bromo derivatives which will consist of two racemic modifications in unequal amounts, and the mixture will be optically inactive.

§7c. Compounds containing three different asymmetric carbon atoms. A molecule of this type is *Cabd·Cab·Cabe*, *e.g.*, the pentoses, and the number of optical isomers possible is *eight* (four pairs of enantiomorphs):









All the cases discussed so far are examples of a series of compounds which contain *n structurally distinct* carbon atoms, *i.e.*, they belong to the series $Cabd{\cdot}(Cab)_{n-2}{\cdot}Cabe$. In general, if there are *n* asymmetric carbon atoms in the molecule (of this series), then there will be 2^n optically active forms and 2^{n-1} resolvable forms (*i.e.*, 2^{n-1} pairs of enantiomorphs). These formulae also apply to *monocyclic* compounds containing *n* different asymmetric carbon atoms; they may or may not apply to *fused ring systems* since spatial factors may play a part in the possible existence of various configurations (see, *e.g.*, camphor, §23a. VIII).

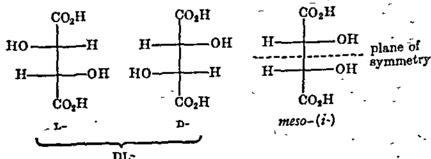
§7d. Compounds of the type $Cabd{\cdot}(Cab)_x{\cdot}Cabd$. In compounds of this type the two *terminal* asymmetric carbon atoms are *similar*, and the number of optically active forms possible depends on whether x is *odd* or *even*.

(i) EVEN SERIES

(*a*) *Cabd·Cabd*, *e g.*, tartaric acid. In a compound of this type the rotatory power of each asymmetric carbon atom is the same. Now let us consider the number of optical isomers possible.

D	L	D	L
D	L	L	D
I	II	III	IV

In molecules I and II, the upper and lower halves reinforce each other; hence I, as a whole, has the dextro- and II, the lævo-configuration, *i.e.*, I and II are optically active, and enantiomorphous. On the other hand, in III the two halves are in opposition, and so the molecule, *as a whole*, will not show optical activity. It is also obvious that III and IV are identical, *i.e.*, there is only *one* optically inactive form of *Cabd·Cabd*. Molecule III is said to be **optically inactive by internal compensation**. Molecule III is known as the **meso** form, and is a diastereoisomer of the pair of enantiomorphs I and II. The *meso* form is also known as the *inactive* form and is represented as the *i*-form; **the meso form cannot be resolved** (see also §10). Thus there are four forms possible for the molecule *Cabd·Cabd*: one pair of enantiomorphs, one racemic modification, and one *meso* (*i*-) form. These forms for tartaric acid are:

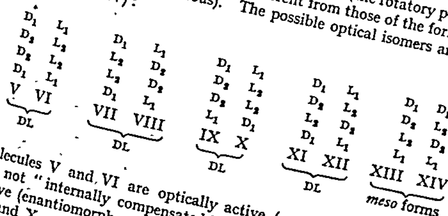


Inspection of these formulae shows that the D- and L- forms do not possess any elements of symmetry; the *meso* form, however, possesses a plane of symmetry.

(*b*) *Cabd·Cab·Cab·Cabd*, *e.g.*, saccharic acid,

$CO_2H·CHOH·CHOH·CHOH·CHOH·CO_2H$.

The rotatory powers of the two terminal asymmetric carbon atoms are the same, and so are those of the middle two (the rotatory powers of the latter are almost certainly different from those of the former, equality would be fortuitous). The possible optical isomers are as follows (V–XIV):

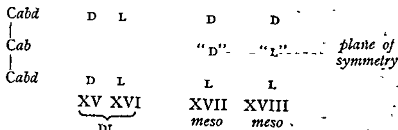


Molecules V and VI are optically active (enantiomorphous) and are not "internally compensated"; VII and VIII are optically active (enantiomorphous) and are not "internally compensated"; IX and X are optically active (enantiomorphous) but are "internally compensated at the ends"; XI and XII are optically active (enantiomorphous) but are "internally compensated in the middle"; XIII and XIV are *meso* forms and are optically inactive by (complete) internal compensation. Thus there are eight optically active forms (four pairs of enantiomorphs), and two *meso* forms.

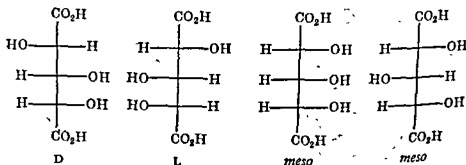
In general, in the series of the type $Cabd·(Cab)_{n-2}·Cabd$, if n is the number of asymmetric carbon atoms and n is *even*, then there will be 2^{n-1} optically active forms, and $2^{\frac{n-2}{2}}$ *meso* forms.

(ii) ODD SERIES

(*a*) *Cabd·Cab·Cabd*, *e.g.*, trihydroxyglutaric acid. If the two terminal asymmetric carbon atoms have the same configuration, then the central carbon atom has two identical groups joined to it and hence cannot be asymmetric. If the two terminal configurations are opposite, then the central carbon atom has apparently four different groups attached to it (the two ends are mirror images and not superimposable). Thus the central carbon atom becomes asymmetric, but at the same time the two terminal atoms "compensate internally" to make the *molecule as a whole* symmetrical (there is now a plane of symmetry), and consequently the compound is not optically active. In this molecule the central carbon

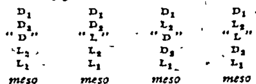


atom is said to be *pseudo-asymmetric*, and is designated "D" and "L" (or ⊕ and ⊖ if the + and − convention is used; §7b). There will, however, be *two meso* forms since the pseudo-asymmetric carbon atom can have two different configurations (see XV–XVIII). Thus there are five forms in all: two optically active forms (enantiomorphs), one racemic modification, and two *meso* forms. The following are the corresponding trihydroxyglutaric acids, all of which are known.

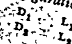


(*b*) *Cabd Cab Cab Cab Cabd*. In this molecule the central carbon atom is pseudo-asymmetric when the left-hand side of the molecule has the opposite configuration to that of the right-hand side; the central carbon atom is symmetrical when both sides have the same configuration. In all other cases the central carbon atom is asymmetric, the molecule now containing five asymmetric carbon atoms. The following table shows that there are *sixteen* optical isomers possible, of which twelve are optically active (six pairs of enantiomorphs), and four are *meso* forms.

Ends with opposite configurations

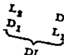


Ends with same configurations





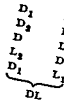




Molecule with five asymmetric carbon atoms









In general, in the series of the type $Cabd{\cdot}(Cab)_{n-2}{\cdot}Cabd$, if n is the number of "asymmetric" carbon atoms and n is *odd*, then there will be 2^{n-1} optical isomers, of which $2^{\frac{n-1}{2}}$ are *meso* forms and the remainder optically active forms.

§8. The racemic modification. The racemic modification is an equimolecular mixture of a pair of enantiomorphs, and it may be prepared in several ways.

(i) Mixing of equimolecular proportions of enantiomorphs produces the racemic modification.

(ii) Synthesis of asymmetric compounds from symmetrical compounds always results in the formation of the racemic modification. This statement is true only if the reaction is carried out in the absence of other optically active compounds or circularly polarised light (see asymmetric synthesis, §7. III).

(iii) **Racemisation.** The process of converting an optically active compound into the racemic modification is known as racemisation. The (+)- and (—)-forms of most compounds are capable of racemisation under the influence of heat, light, or chemical reagents. Which agent is used depends on the nature of the compound, and at the same time the ease of racemisation also depends on the nature of the compound, e.g.

(*a*) Some compounds racemise so easily that they cannot be isolated in the optically active forms.

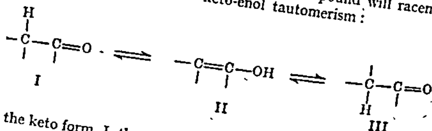
(*b*) A number of compounds racemise spontaneously when isolated in optically active forms.

(*c*) The majority of compounds racemise with various degrees of ease under the influence of different reagents.

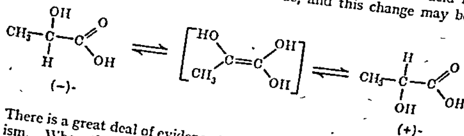
(*d*) A relatively small number of compounds cannot be racemised at all.

Many theories have been proposed to explain racemisation, but owing to the diverse nature of the structures of the various optically active compounds, one cannot expect to find *one* theory which would explain the racemisation of *all types* of optically active compounds. Thus we find that a number of mechanisms have been suggested, each one explaining the racemisation of a particular type of compound.

A number of compounds which are easily racemisable are those in which the asymmetric carbon atom is joined to a hydrogen atom and a negative group. Since this type of compound can undergo tautomeric change, the mechanism proposed for this racemisation is one *via* enolisation. When the intermediate enol form, which is symmetrical, reverts to the keto form, it can do so equally well to produce the (+)- or (−)-forms, *i.e.*, the compound will racemise. Let us consider the case of keto-enol tautomerism:



In the keto form, I, the carbon joined to the hydrogen atom and the oxo group is asymmetric, in the enol form, II, this carbon atom has lost its asymmetry. When the enol form reverts to the keto form, it can do so to produce the original keto molecule I, but owing to its symmetry, the enol form can produce equally well the keto form III in which the configuration of the asymmetric carbon atom is opposite to that in I. Thus racemisation, according to this scheme, occurs *via* the enol form, *e.g.*, (−)-lactic acid is racemised in aqueous sodium hydroxide, and this change may be formulated:



There is a great deal of evidence to support this tautomeric mechanism. When the hydrogen atom joined to the asymmetric carbon atom is replaced by some group that prevents tautomerism (enolisation) then racemisation is also prevented (at least under the same,

conditions as the original compound), *e.g.*, mandelic acid is readily racemised by warming with aqueous sodium hydroxide.

$$\underset{(+)\text{-}}{C_6H_5\text{—}C(H)(OH)\text{—}C(=O)OH} \rightleftharpoons \left[\begin{matrix} C_6H_5 \\ HO \end{matrix}\!>C{=}C<\!\begin{matrix} OH \\ OH \end{matrix}\right] \rightleftharpoons \underset{(-)\text{-}}{C_6H_5\text{—}C(OH)(H)\text{—}C(=O)OH}$$

On the other hand, atrolactic acid, $C_6H_5{\cdot}C(CH_3)(OH){\cdot}CO_2H$, is not racemised under the same conditions; in this case keto-enol tautomerism is no longer possible.

Racemisation of compounds capable of exhibiting keto-enol tautomerism is catalysed by acids and bases. Since keto-enol tautomerism is also catalysed by acids and bases, then if racemisation proceeds *via* enolisation, the rates of racemisation and enolisation should be the same. This relationship has been established by means of kinetic studies, *e.g.*, Bartlett *et al.* (1935) found that the rate of acid-catalysed iodination of 2-butyl phenyl ketone was the same as that of racemisation in acid solution. This is in keeping with both reactions involving the rate-controlling formation of the enol (see Vol. I, Ch. X):

$$\underset{(+)}{Ph{\cdot}CO{\cdot}CHMeEt} \underset{fast}{\overset{slow}{\rightleftharpoons}} Ph{\cdot}C(OH){=}CMeEt \underset{slow}{\overset{fast}{\rightleftharpoons}} \underset{(-)}{Ph{\cdot}CO{\cdot}CHMeEt}$$

$$Ph{\cdot}C(OH){=}CMeEt \xrightarrow{I_2,\ fast} Ph{\cdot}CO{\cdot}CIMeEt$$

On the other hand, on the basis that the rate-determining step in base-catalysed enolisation and racemisation is the formation of the enolate ion, then the two processes will also occur at the same rate.

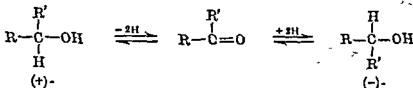
$$B + R{\cdot}CO{\cdot}CHR_2 \xrightarrow{slow} BH^+ + R{\cdot}C(O^-){=}CR_2 \xrightarrow{fast} B + R{\cdot}C(OH){=}CR_2$$

Hsü *et al.* (1936) found that the rates of bromination and racemisation (in the presence of acetate ions) of 2-*o*-carboxybenzyl-1-indanone were identical.

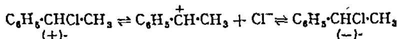
$$\text{(indanone ring: } C_6H_4\text{—}CH_2\text{—}CH(\text{—}CH_2\text{—}C_6H_4{\cdot}CO_2H)\text{—}CO\text{)}$$

There are many compounds containing an asymmetric carbon

atom which can be racemised under suitable conditions although there is no possibility of tautomerism. A number of different types of compounds fall into this group, and the mechanism proposed for racemisation depends on the type of compound under consideration. In the case of compounds of the type of (−)-limonene (§13. VIII), which is racemised by strong heating, the mechanisms proposed are highly speculative (see, for example, Werner's theory, §4. V). A number of optically active secondary alcohols can be racemised by heating with a sodium alkoxide. This has been explained by a reversible dehydrogenation (Hückel, 1931) and there is some evidence to support this mechanism (Doering *et al.*, 1947, 1949).

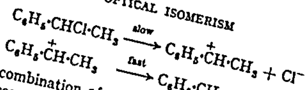


Another different type of compound which can be readily racemised is that represented by α-chloroethylbenzene. When the (+)- or (−)-form is dissolved in liquid sulphur dioxide, spontaneous racemisation occurs. This has been explained by assuming ionisation into a carbonium ion (Polanyi *et al.*, 1933).



The carbonium ion is planar (the positively charged carbon atom is probably in a state of trigonal hybridisation) and consequently symmetrical; recombination with the chlorine ion can occur equally well to form the (+)- and (−)-forms, *i.e.*, racemisation occurs. The basis of this mechanism is that alkyl halides in liquid sulphur dioxide exhibit an electrical conductivity, which has been taken as indicating ionisation. Hughes, Ingold *et al.* (1936), however, found that pure α-chloroethylbenzene in pure liquid sulphur dioxide does not conduct, but when there is conduction, then styrene and hydrogen chloride are present. These authors showed that under the conditions of purity, the addition of bromine leads to a quantitative yield of styrene dibromide.

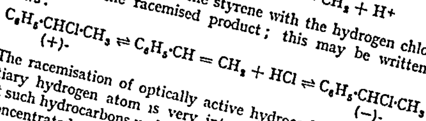
Polanyi showed that the rate of racemisation of α-chloroethylbenzene in liquid sulphur dioxide is unaffected by added chloride ions. Hughes and Ingold suggest that the rate of racemisation is accounted for by the rate of formation of hydrogen chloride; thus:



$$C_6H_5{\cdot}CHCl{\cdot}CH_3 \xrightarrow{\text{slow}} C_6H_5{\cdot}\overset{+}{C}H{\cdot}CH_3 + Cl^-$$

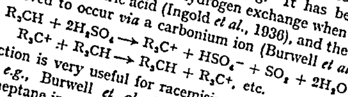
$$C_6H_5{\cdot}\overset{+}{C}H{\cdot}CH_3 \xrightarrow{\text{fast}} C_6H_5{\cdot}CH = CH_2 + H^+$$

It is the recombination of the styrene with the hydrogen chloride that produces the racemised product; this may be written as follows:



$$\underset{(+)\text{-}}{C_6H_5{\cdot}CHCl{\cdot}CH_3} \rightleftharpoons C_6H_5{\cdot}CH = CH_2 + HCl \rightleftharpoons \underset{(-)\text{-}}{C_6H_5{\cdot}CHCl{\cdot}CH_3}$$

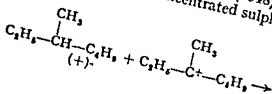
The racemisation of optically active hydrocarbons containing a tertiary hydrogen atom is very interesting. It has been shown that such hydrocarbons undergo hydrogen exchange when dissolved in concentrated sulphuric acid (Ingold *et al.*, 1936), and the mechanism is believed to occur *via* a carbonium ion (Burwell *et al.*, 1948).



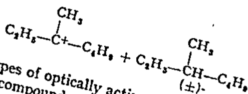
$$R_3CH + 2H_2SO_4 \longrightarrow R_3C^+ + HSO_4^- + SO_2 + 2H_2O$$

$$R_3C^+ + R_3CH \longrightarrow R_3CH + R_3C^+, \text{ etc.}$$

This reaction is very useful for racemising optically active hydrocarbons, *e.g.*, Burwell *et al.* (1948) racemised optically active 3-methylheptane in concentrated sulphuric acid (the carbonium ion is flat):



$$\underset{(+)\text{-}}{C_2H_5{-}CH(CH_3){-}C_4H_9} + C_2H_5{-}\overset{+}{C}(CH_3){-}C_4H_9 \longrightarrow$$



$$C_2H_5{-}\overset{+}{C}(CH_3){-}C_4H_9 + \underset{(\pm)\text{-}}{C_2H_5{-}CH(CH_3){-}C_4H_9}$$

The racemisation of other types of optically active compounds is described later (see diphenyl compounds, §4. V; nitrogen compounds, §2a. VI; phosphorus compounds, §3b. VI; arsenic compounds, §4a. VI).

§9. Properties of the racemic modification. The racemic modification may exist in three different forms in the solid state.

(i) **Racemic mixture.** This is also known as a (±)-**conglomerate**, and is a mechanical mixture of two types of crystals, the (+)- and (−)-forms; there are two phases present.

(ii) **Racemic compound.** This consists of a pair of enantiomorphs in combination as a molecular compound; only one solid phase is present. The physical properties of a racemic compound are different from those of the constituent enantiomorphs, but in solution racemic compounds dissociate into the (+)- and (−)-forms.

(iii) **Racemic solid solution.** This is also known as a ***pseudo*-racemic compound**, and is a solid solution (one phase system) formed by a pair of enantiomorphs crystallising together due to their being isomorphous.

§9a. Methods for determining the nature of a racemic modification. One simple method of examination is to estimate the amounts of water of crystallisation in the enantiomorphs (only one need be examined) and in the racemic modification; if these are different, then the racemic modification is a racemic compound. Another simple method is to measure the densities of the enantiomorphs and the racemic modification; again, if these are different, the racemic modification is a racemic compound; *e.g.*, the tartaric acids.

	D-Tartaric acid	L-Tartaric acid	*Racemic* Tartaric acid
Melting point . . .	170°	170°	206°
Water of crystallisation	None	None	$1H_2O$
Density	1·7598	1·7598	1·697
Solubility in H_2O (at 20°)	139 g./100 ml.	139 g /100 ml.	20·6 g./100 ml.

There are, however, two main methods for determining the nature of a racemic modification: a study of the freezing-point curves and a study of the solubility curves (Roozeboom, 1899; Andriani, 1900).

Freezing-point curves. These are obtained by measuring the melting points of mixtures containing different amounts of the

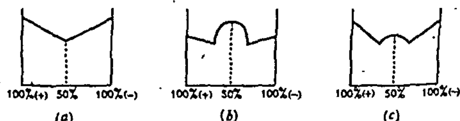


Fig. 2 18.

[illegible] and its corresponding enantiomorphs. Various [illegible] racemic [illegible] ures are higher than that of the racemic modification alone. In this case the racemic modification is a racemic mixture (a eutectic mixture is formed at the point of 60 per cent. composition of each enantio-

morph), and so addition of either enantiomorph to a racemic mixture *raises* the melting point of the latter; (±)-pinene is an example of this type. In Fig. 18 (*b*) and (*c*) the melting points of the mixtures are lower than the melting point of the racemic modification which, therefore, is a racemic compound. The melting point of the racemic compound may be above that of each enantiomorph (Fig. 18 *b*) or below (Fig. 18 *c*); in either case the melting point is *lowered* when the racemic compound is mixed with an enantiomorph; an example of Fig. 18 (*b*) is methyl tartrate, and one of Fig. 18 (*c*) is mandelic acid.

When the racemic modification is a racemic solid solution, three types of curves are possible (Fig. 19). In Fig. 19 (*a*) the freezing-point curve is a horizontal straight line, all possible compositions having the same melting point, *e.g.*, (+)- and (−)- camphor. In Fig. 19 (*b*) the freezing-point curve shows a maximum, *e.g.*, (+)- and (−)-carvoxime; and in Fig. 19 (*c*) the freezing-point curve shows a minimum, *e.g.*, (+)- and (−)-*iso*pentyl (*iso*amyl) carbamate.

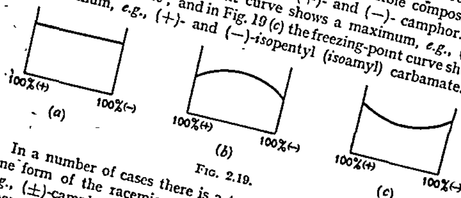
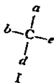
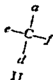


FIG. 2.19.

In a number of cases there is a transition temperature at which one form of the racemic modification changes into another form, *e.g.*, (±)-camphoroxime crystallises as the racemic solid solution above 103°, whereas below this temperature it is the racemic compound that is obtained [see also §10(*i*)].

Fredga (1944) has introduced the study of quasi-racemic compounds as a means of correlating configurations (§5). Quasi-racemic compounds are equimolecular compounds that are formed from two optically active compounds which have *closely similar structures but opposite configurations*, *e.g.*, I and II. The formation





of a quasi-racemic compound is detected by studying the melting-point curves of the two components. The curves obtained are

similar to those of the racemic modification shown in Fig. 18 (*a*), 18 (*b*), and 19 (*a*), but with the quasi-racemic compounds these curves are unsymmetrical (since the m.p.s. of the components will be different). An unsymmetrical curve 18 (*a*) indicates a eutectic mixture, an unsymmetrical 19 (*a*) a *solid solution*, and an unsymmetrical 18 (*b*) a quasi-racemic compound. Curves for quasi-racemic compounds are given only by compounds (containing one asymmetric carbon atom) which have closely similar structures but opposite configurations. On the other hand, curves of the other two types are given by compounds of *like* configuration (but some cases are known where the configurations have been opposite). Various examples of this method of correlating configurations have now been described, *e.g.*, (+)-malic acid and (+)-methylsuccinic acid have the same configuration since these two compounds form a quasi-racemic compound (see also §§10(vi) and 23e. VIII).

Solubility curves. The interpretation of solubility curves is difficult, but in practice the following simple scheme based on solubility may be used. A small amount of one of the enantiomorphs is added to a *saturated* solution of the racemic modification, and the resulting solution is then examined in a polarimeter. If the solution exhibits a rotation, then the racemic modification is a compound, but if the solution has a zero rotation, then the racemic modification is a mixture or a solid solution. The reasons for this behaviour are as follows. If the racemic modification is a mixture or a solid solution, then the solution (in some solvent) is saturated with respect to each enantiomorph and consequently cannot dissolve any of the added enantiomorph. If, however, the racemic modification is a compound, then the solution (in a solvent) is saturated with respect to the compound form but not with respect to either enantiomorph; hence the latter will dissolve when added and thereby produce a rotation. It should be noted that this simple method does not permit a differentiation to be made between a racemic mixture and a racemic solid solution.

§10. **Resolution of racemic modifications.** *Resolution is the* process whereby a racemic modification is separated into its two enantiomorphs. In practice the separation may be far from quantitative, and in some cases only one form may be obtained. A large variety of methods for resolution have now been developed, and the method used in a particular case depends largely on the chemical nature of the compound under consideration.

(i) **Mechanical separation.** This method is also known as **spontaneous resolution**, and was introduced by Pasteur (1848). It depends on the crystallisation of the two forms separately.

which are then separated by hand. The method is applicable only to a few cases, and then only for racemic *mixtures* where the *crystal* forms of the enantiomorphs are themselves enantiomorphous (§2). Pasteur separated sodium ammonium racemate in this way. The transition temperature of sodium ammonium racemate is 28°; above this temperature the racemic compound crystallises out, and below this temperature the racemic mixture. Now Pasteur crystallised his sodium ammonium racemate from a concentrated solution at room temperature, which must have been below 28° since had the temperature been above this he would have obtained the racemic compound, which cannot be separated mechanically. Actually, Staedel (1878) failed to repeat Pasteur's separation since he worked at a temperature above 28°.

(ii) **Preferential crystallisation by inoculation.** A supersaturated solution of the racemic modification is treated with a crystal of one enantiomorph (or an isomorphous substance), whereupon this form is precipitated.

(iii) **Biochemical separation** (Pasteur, 1858). Certain bacteria and moulds, when they grow in a dilute solution of a racemic modification, destroy one enantiomorph more rapidly than the other, *e.g.*, *Penicillium glaucum* (a mould), when grown in a solution of ammonium racemate, attacks the D-form and leaves the L-.

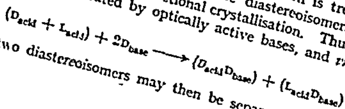
This biochemical method of separation has some disadvantages:

(*a*) Dilute solutions must be used, and so the amounts obtained will be small.

(*b*) One form is always destroyed and the other form is not always obtained in 50 per cent. yield since some of this may also be destroyed.

(*c*) It is necessary to find a micro-organism which will attack only one of the enantiomorphs.

(iv) **Conversion into diastereoisomers** (Pasteur, 1858). This method, which is the best of all the methods of resolution, consists in converting the enantiomorphs of a racemic modification into diastereoisomers (§7b); the racemic modification is treated with an optically active substance and the diastereoisomers thereby produced are separated by fractional crystallisation. Thus racemic acids may be separated by optically active bases, and *vice versa*, *e.g.*,



$$(D_{acid} + L_{acid}) + 2D_{base} \longrightarrow (D_{acid}D_{base}) + (L_{acid}D_{base})$$

These two diastereoisomers may then be separated by fractional

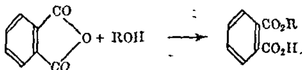
crystallisation and the acids (enantiomorphs) regenerated by hydrolysis with inorganic acids or with alkalis. In practice it is usually easy to obtain the less-soluble isomer in a pure state, but it may be very difficult to obtain the more-soluble isomer. In a number of cases this second (more-soluble) isomer may be obtained by preparing it in the form of *another* disastereoisomer which is less soluble than that of its enantiomorph.

Resolution by means of diastereoisomer formation may be used for a variety of compounds, *e.g.*,

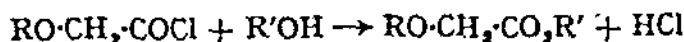
(*a*) *Acids.* The optically active bases used are mainly alkaloids: brucine, quinine, strychnine, cinchonine, cinchonidine and morphine. Recently, optically active benzimınazoles (§3a. XII) have been used (Hudson *et al.*, 1939).

(*b*) *Bases.* Many optically active acids have been used, *e.g.*, tartaric acid, camphor-β-sulphonic acid, and particularly α-bromocamphor-π-sulphonic acid (see §23a. VIII).

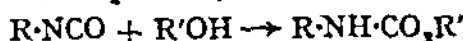
(*c*) *Alcohols.* These are converted into the acid ester derivative using either succinic or phthalic anhydride (Pickard and Kenyon, 1912). The acid ester, consisting of equimolecular amounts of the



(+)- and (−)-forms, may now be resolved as for acids. Racemic alcohols may also be resolved by diastereoisomer formation with optically active acyl chlorides (to form esters) or with optically active *iso*cyanates (to form urethans):

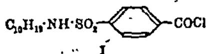


$$RO{\cdot}CH_2{\cdot}COCl + R'OH \rightarrow RO{\cdot}CH_2{\cdot}CO_2R' + HCl$$



$$R{\cdot}NCO + R'OH \rightarrow R{\cdot}NH{\cdot}CO_2R'$$

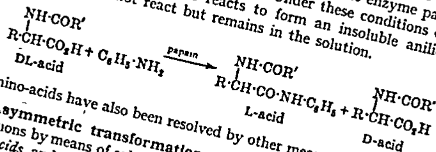
In these equations R is the (−)-menthyl radical (§16. VIII); recently *N*-(−)-menthyl-*p*-sulphamylbenzoyl chloride, I, has been used (Mills *et al.*, 1950).



(*d*) *Aldehydes and Ketones.* These have been resolved by means of optically active hydrazines, *e.g.*, (−)-menthylhydrazine. Sugars have been resolved with (+)-*iso*pentanethiol (*cf.* §1. VII). Nerdel *et al.* (1952) have resolved oxo compounds with D-tartramide acid

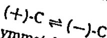
hydrazide, $NH_2 \cdot CO \cdot CHOH \cdot CHOH \cdot CO \cdot NH \cdot NH_2$; this forms diastereoisomeric tartramazones.

(*e*) *Amino-compounds.* These may be resolved by conversion into diastereoisomeric anils by means of optically active aldehydes. α-Amino-acids have been resolved by preparing the acyl derivative with an optically active acyl chloride, *e.g.*, (−)-menthoxyacetyl chloride (*cf. alcohols*). Another method of resolving DL-amino-acids is asymmetric enzymic synthesis (§7. III). The racemic amino-acid is converted into the acyl derivative which is then allowed to react with aniline in the presence of the enzyme papain at the proper *p*H (Albertson, 1951). Under these conditions only the L-amino-acid derivative reacts to form an insoluble anilide; the D-acid does not react but remains in the solution.



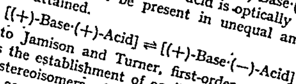
Amino-acids have also been resolved by other means (see §4. XIII).

Asymmetric transformation. Resolution of racemic modifications by means of salt formation (the diastereoisomers are salts; *cf. acids* and *bases*) may be complicated by the phenomenon of *asymmetric transformation*. This phenomenon is exhibited by compounds that are optically unstable, *i.e.*, the enantiomorphs are readily interconvertible:



There are two types of asymmetric transformation, first order and second order. These were originally defined by Kuhn (1932), but were later re-defined by Jamison and Turner (1942).

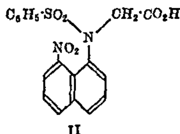
Suppose we have an optically stable (+)-base (one equivalent) dissolved in some solvent, and this is then treated with one equivalent of an optically unstable (±)-acid. At the moment of mixing the solution will contain equal amounts of [(+)-Base·(+)-Acid] and [(+)-Base·(−)-Acid]; but since the acid is optically unstable, the two diastereoisomers will be present in unequal amounts when equilibrium is attained.



According to Jamison and Turner, first-order asymmetric transformation is the establishment of equilibrium *in solution* between the two diastereoisomers which must have a *real* existence. In second-order asymmetric transformation it is necessary that one salt should crystallise from solution; the two diastereoisomers

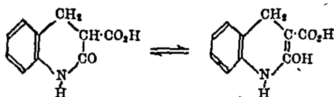
need not have a real existence in solution. In second-order asymmetric transformation it is possible to get a complete conversion of the acid into the form that crystallises; the form may be the (+)- or (−)-, and which one it is depends on the nature of the base and the solvent.

Many examples of first- and second-order asymmetric transformation are known, and a large number of these compounds are those which owe their asymmetry to restricted rotation about a single bond (see Ch. V), *e.g.*, Mills and Elliott (1928) tried to resolve *N*-benzenesulphonyl-8-nitro-1-naphthylglycine, II, by means of the



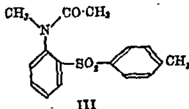
II

brucine salt. These authors found that either diastereoisomer could be obtained in approximately 100 per cent. yield by crystallisation from methanol and acetone, respectively. Another example of second-order asymmetric transformation is hydrocarbostyril-3-carboxylic acid. This compound contains an asymmetric carbon



atom, and Leuchs (1921), attempting to resolve it with quinidine, isolated approximately 90 per cent. of the (+)-form. Optical instability in this case is due to keto-enol tautomerism (*cf.* §8).

A very interesting example of second-order asymmetric transformation is 2-acetomethylamido-4′:5-dimethylphenylsulphone, III.



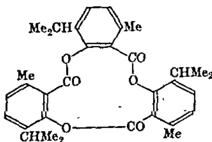
III

(+)-mandelate than (−)-menthyl (−)-mandelate. Consequently there will be more (−)-mandelic acid than (+)-mandelic acid in the *unchanged* acid, *i.e.*, a partial resolution of (±)-mandelic acid has been effected (see also §5b. VI).

(viii) Ferreira (1953) has partially resolved (±)-narcotine and (±)-laudanosine (1–2·5 per cent. resolution) *without* the use of optically active reagents. He dissolved the racemic alkaloid in hydrochloric acid and then *slowly* added pyridine; the alkaloid was precipitated, and it was found to be optically active. The explanation offered for this partial resolution is as follows (Ferreira). When a crystalline racemic substance is precipitated from solution, a crystallisation nucleus is first developed. Since this nucleus contains a relatively small number of molecules, there is more than an even chance that it will contain an excess of one enantiomorph or other. If it be assumed that the forces acting on the growth of crystals are the same kind as those responsible for adsorption [*cf.* (vi)], the nucleus will grow preferentially, collecting one enantiomorph rather than the other. Crystallisation, when carried out in the usual manner, results in the formation of crystals containing more or less equivalent numbers of both enantiomorphs.

Channel complex formation has also been used to resolve racemic modifications (see Vol. I). This also offers a means of carrying out a resolution without asymmetric reagents, *e.g.*, Schlenk (1952) added (±)-2-chloro-octane to a solution of urea and obtained, on fractional crystallisation, the two urea inclusion complexes urea/(+)-2-chloro-octane and urea/(−)-2-chloro-octane.

Baker *et al.* (1952) have prepared tri-*o*-thymotide, and found that it formed clathrates with ethanol, *n*-hexane, etc. Powell *et al.* (1952) have shown that tri-*o*-thymotide crystallises as a racemate, but that resolution takes place when it forms clathrates with *n*-hexane, benzene or chloroform. By means of seeding and slow growth of a single crystal, it is possible to obtain the (+)- or (−)-form depending on the nature of the seed. Furthermore, crystallisation of tri-*o*-thymotide (*dl*) from a solvent which is itself a racemic modification (*d′l′*) and which forms a clathrate, produces crystals of the types *dd′* and *ll′*. Thus such (solvent) racemic modifications can be resolved, *e.g.*, *sec.*-butyl bromide has been resolved in this way.



tri-*o*-thymotide

than that for the dextro component, the substance will be dextrorotatory. The difficulty of Fresnel's theory is that it does not

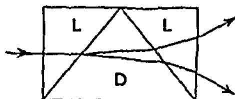


Fig. 2.21.

explain *why* the two circular components should travel with different, velocities. It is interesting to note, however, that Fresnel (1824) suggested that the optical activity of quartz is due to the structure being built-up in right- and left-handed spirals (*cf.* §2).

Now let us consider the problem of optical activity of substances *in solution*. In this case the optical activity is due to the *molecules* themselves, and not to crystalline structure (see also §2). Any *crystal* which has a plane of symmetry *but not a centre of symmetry* (§6) rotates the plane of polarisation, the rotation varying with the direction in which the light travels through the crystal. No rotation occurs if the direction of the light is perpendicular or parallel to the plane of symmetry. If we assume that molecules in a solution (or in a pure liquid) behave as individual crystals, then any molecule having a plane but not a centre of symmetry will also rotate the plane of polarisation, provided that the light travels through the molecule in any direction other than perpendicular (or parallel) to the plane of symmetry. Let us consider the molecule Ca_2bd (Fig. 22). This has a plane of symmetry, and so molecule I and its mirror image II are superimposable. Now let us suppose that the direction of plane-polarised light passing through molecule I makes an angle $\theta°$ with the plane of symmetry, and that the resultant rotation is $+\alpha°$. Then if the direction of the light through molecule II also makes an angle $\theta°$ with the plane of symmetry, the resultant rotation will be $-\alpha°$. Thus the *total* rotation produced by molecules I and II is *zero*. In a solution of compound Ca_2bd, there will be an *infinite number of molecules in random orientation*. Statistically one can expect to find that whatever the angle θ is for molecule I, there will always be molecule II also being traversed by light entering at angle θ. **Thus, although each individual molecule rotates the plane of polarisation by an amount depending on the value of θ, the statistical sum of the contributions of the individual molecules will be zero.**

When a molecule is not superimposable on its mirror image,

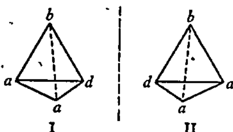


FIG. 2.22.

then if only one enantiomorph is present in the solution, the rotation produced by each individual molecule will (presumably) depend on the angle of incidence (with respect to any face), but there will be no compensating molecules (*i.e.*, mirror image molecules) present. Hence, in this case, there will be a net rotation that is *not* zero, the actual value being the statistical sum of the individual contributions (which are all in the *same* direction). Thus, if we consider the behaviour of a compound in a solution (or as a pure liquid) *as a whole*, then the observed experimental results are always in accord with the statement that **if the molecular structure of the compound is asymmetric, that compound will be optically active** (§2). Any compound composed of molecules possessing a plane but not a centre of symmetry is, considered *as a whole*, optically inactive, the net zero rotation being the result of "external compensation" (*cf.* §7a). This point is of great interest in connection with flexible molecules (§4). Let us consider *meso*tartaric acid, a compound that is optically inactive by internal compensation (§7b). X-ray studies (Stern *et al.*, 1950) have shown that the *trans* position of the molecule is the favoured one (Fig. 23 *a*). This has a centre of symmetry, and so molecules in this configuration are *individually* optically inactive. On the

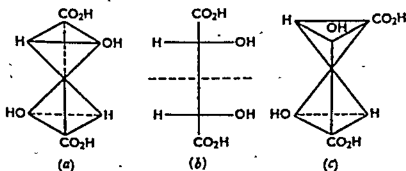


FIG. 2.23.

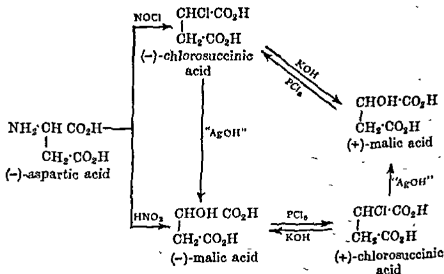
van't Hoff, *Chemistry in Space*, Oxford Press (1891, translated by Marsh).
Bijvoet, Structure of Optically Active Compounds in the Solid State, *Nature*, 1954, **173**, 888.
Powell, New Procedure for Resolution of Racemic Substances, *Nature*, 1952, **170**, 155.
Rosanoff, On Fischer's Classification of Stereoisomers, *J. Amer Chem Soc.*, 1906, **28**, 114.
Turner and Harris, Asymmetric Transformation and Asymmetric Induction, *Quart. Reviews (Chem. Soc)*, 1948, **1**, 299
Arnstein and Bentley, Isotopic Tracer Technique, *Quart Reviews (Chem Soc.)*, 1950, **4**, 191.
Bent, Aspects of Isomerism and Mesomerism, *J. Chem. Educ.*, 1953, **30**, 220, 284, 328.
Kauzmann, Walter and Eyring, Theories of Optical Rotatory Power, *Chem. Reviews*, 1940, **26**, 339.
Hudson, Emil Fischer's Stereo-Formulas, *Advances in Carbohydrate Chemistry*, Academic Press. Vol 3 (1948). Ch. I
Ferreira, Resolution of Racemic Mixtures by Symmetrical Reagents, *Nature*, 1953, **171**, 39.
McCasland and Proskow, A Compound whose Molecule is superposable on its Mirror Image but contains no Plane or Centre of Symmetry, *J. Amer. Chem. Soc.*, 1955, **77**, 4689
Barton and Cookson, The Principles of Conformational Analysis, *Quart Reviews (Chem Soc.)*, 1956, **10**, 44.
Newman (Ed.), *Steric Effects in Organic Chemistry*, Wiley (1956). Ch I. Conformational Analysis.
Mizushima, *Structure of Molecules and Internal Rotation*, Academic Press (1954).
Whiffen, Optical Rotation and Geometrical Structure, *Chem. and Ind.*, 1956, 964.
Truter, Sorting Molecules by Size and Shape, *Research*, 1953, **6**, 320.
Klyne (Ed.), *Progress in Stereochemistry*, Butterworth Vol. I (1954); Vol. II (1958).

CHAPTER III

WALDEN INVERSION

§1. Nature of the Walden inversion (Optical inversion). By means of a series of replacement reactions, Walden (1893) was able to transform an optically active compound into its enantiomorph. In some cases the product is 100 per cent. optically active, *i.e.*, the *inversion* is quantitative; in other cases the product is a mixture of the (+)- and (−)-forms, but in unequal amounts, *i.e.*, a partial inversion has taken place.

The phenomenon was first discovered by Walden during an investigation of the properties of (−)-aspartic acid. The following chart shows the reactions carried out by Walden.



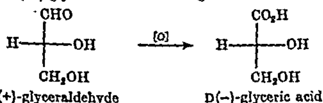
The change:

(−)-malic acid I $\xrightarrow{PCl_5}$ (+)-chlorosuccinic acid II $\xrightarrow{\text{"AgOH"}}$ (+)-malic acid III

constitutes a **Walden inversion**, which may be defined as the conversion of the (+)-form into the (−)-form, or *vice versa*, without recourse to resolution. In one, and only one, of the two reactions there must be an interchange of position between the two groups, *e.g.*, if the configuration of I corresponds with that of II, the inversion of configuration must have taken place between II and III.

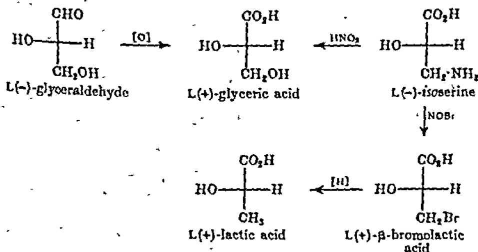
The "definition" of the Walden inversion given above is the one that was used by Fischer (1906). The tendency nowadays is to apply the term *Walden inversion* to any *single* reaction in which inversion of configuration takes place.

As already pointed out, in the conversion of (−)-malic acid into (+)-malic acid *via* (+)-chlorosuccinic acid, inversion occurs in only one of the two reactions; if inversion occurred in both, then the original (−)-malic acid will be produced. As the experiment stands, there is no way of telling which stage is accompanied by inversion; **change in sign of rotation does not necessarily mean that inversion of configuration has occurred.** A simple example that illustrates this point is the oxidation of D(+)-glyceraldehyde to D(−)-glyceric acid. The sign of rotation has changed,



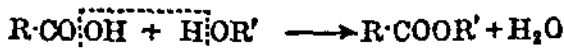
D(+)-glyceraldehyde D(−)-glyceric acid

but the asymmetric carbon atom is not affected by the reaction, *i.e.*, both compounds have the *same relative configurations.* Thus it is necessary to have a classification of relative configurations. The scheme in practice is to use D(+)-glyceraldehyde as arbitrary standard, and to build up from this the D- and L-series of compounds (see §5. II). In some cases it is fairly easy to determine relative configurations, *e.g.*, D(+)-glyceraldehyde and D(−)-glyceric acid. Another example is L(+)-lactic acid, the relative configuration of which is shown in the following series of reactions (Freudenberg, 1914).

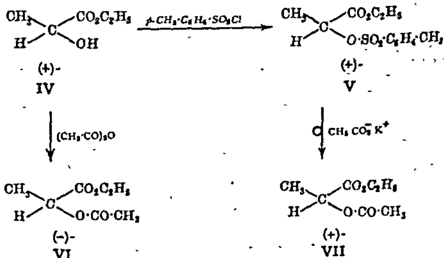


L(−)-glyceraldehyde L(+)-glyceric acid L(−)-isoserine L(+)-lactic acid L(+)-β-bromolactic acid

In all of these reactions, the asymmetric carbon atom is not affected, and consequently no inversion can occur. When, however, the asymmetric carbon atom is involved in the various reactions, the problem of ascertaining relative configurations becomes more difficult owing to the possibility of the Walden inversion occurring at one or more stages. Kenyon and Phillips (1925) established a basis for the determination of relative configurations by the following chemical methods. These authors carried out a series of reactions on optically active hydroxy compounds. Now it has been established that in esterification of an alcohol by a monocarboxylic acid under ordinary conditions, the oxygen atom of the alcoholic group is retained in the ester (see Vol. I, Ch. IX):



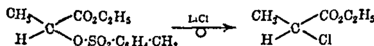
Kenyon and Phillips assumed that in all reactions of this type, the R—O bond of the alcohol remained intact and consequently *no inversion of the alcohol configuration is possible*. Let us now consider the following chart which shows a series of reactions carried out by Kenyon and Phillips on ethyl (+)-lactate; a conventional way of indicating inversion at a particular stage is to use the symbol ⟶.



Ester VI has the same configuration as the parent substance IV, although the sign of rotation has changed; similarly, ester V has the same configuration as IV. Reaction of V (the *p*-toluenesulphonate) with potassium acetate produces ester VII, which is the enantiomorph of VI. Therefore inversion must have occurred

at this stage (on the assumption that V and VI are produced *without* inversion for the reasons given above). It is important to note that if inversion is going to take place at all, the *complete group* attached to an asymmetric carbon atom must be removed. The converse, however, is not necessarily true, *i.e.*, removal of a complete group from an asymmetric carbon atom does not invariably result in inversion (see §4).

The above series of reactions has been used as a standard, and all closely analogous reactions are assumed to behave in a similar way, *i.e.*, inversion and retention of configuration occur at the same stages as those in the standard reaction; *e.g.*, the action of lithium chloride on the *p*-toluenesulphonate is assumed to be analogous to that of potassium acetate, and so the chloride produced has an inverted configuration.

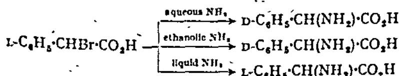


By similar procedures, Kenyon *et al* (1929, 1930) showed that (+)-octan-2-ol and (+)-2-chloro, -2-bromo- and -2-iodo-octane have the same relative configurations; and also that (+)-α-hydroxyethylbenzene, $C_6H_5·CHOH·CH_3$, (+)-α-chloro- and α-bromoethylbenzene have the same relative configurations.

§2. Factors affecting the Walden inversion. As a result of a large amount of work on the Walden inversion, it has been found that at least three factors play a part in deciding whether inversion or retention of configuration will occur.

(i) *Nature of the reagent.* The replacement of halogen by hydroxyl is usually accompanied by inversion when the reagent used is sodium or potassium hydroxide. On the other hand, when silver hydroxide (moist silver oxide) is used, there is usually retention of configuration. Replacement of hydroxyl by chlorine is usually accompanied by inversion when the reagents used are phosphorus tri- and pentachloride, hydrogen chloride, and thionyl chloride.

(ii) *Nature of the solvent.* A number of cases are known where the nature of the solvent affects the steric course of the reaction, *e.g.*, the conversion of L-(α-bromo)phenylacetic acid into (α-amino)-phenylacetic acid by means of ammonia (Senter *et al.*, 1915).



where k is the specific reaction rate, A is the *frequency factor*, and E the *energy of activation* and represents the minimum amount of energy that molecules must acquire before they can react on collision.

The most direct method of calculating the velocity of a bimolecular reaction in the gas phase is to calculate the number of collisions, Z, which the reactant molecules (contained in a unit volume) will make per second. The value of Z was found to be far too great, *i.e.*, the actual rate was slower than the theoretical value based on reaction occurring on each collision. Thus all collisions are not effective. As we have seen, only those molecules with the necessary energy of activation will react; thus we may write the Arrhenius equation as:

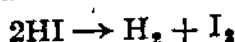
$$k = Ze^{-\frac{E}{RT}}$$

It was found that the specific reaction rates of many bimolecular reactions in the gas phase agreed with this equation within a factor of 10. On the other hand, many bimolecular gas reactions did not agree with this equation, the reaction being much slower than expected. Thus a factor P was introduced:

$$k = PZe^{-\frac{E}{RT}}$$

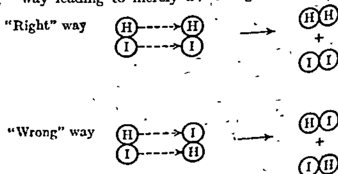
P is known as the *probability* or *steric factor*, and its value is a measure of the deviation of the actual reaction from that calculated on the collision theory. The values of P are often between 10^{-1} to 10^{-3}, but much lower values have also been observed.

A simple example of the steric factor is that in the reaction



$$2HI \longrightarrow H_2 + I_2$$

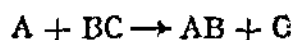
If hydrogen iodide decomposes on collision, then molecules possessing the requisite amount of energy of activation can collide in one of two ways, the "right" way leading to decomposition, and the "wrong" way leading to merely a "change in partners".



Now let us consider reactions in solution. It has been found that, in general, the rate of reaction for bimolecular reactions in which the reactants and products are of *low* polarity is of the same order in the gas phase and in solution. Furthermore, the rate for reactions in solution is not very much affected by the nature of the solvent. If, however, the reactant molecules are polar, then the reaction rate in solution is different from that in the gas phase. In solution, the forces between the reactant molecules will be altered according to the nature of the solvent, particularly its dielectric constant. The values of E, P and Z are influenced by the solvent, and the effects are greatest for reactions involving molecules with a permanent dipole moment, or ions produced from the reacting molecules by the solvent. In general, reactions in which the products are more polar than the reactants are favoured by polar solvents, whereas reactions in which the products are less polar than the reactants are favoured by non-polar solvents. If the reactant molecules are solvated, the reaction rate is decreased because of the increase in E.

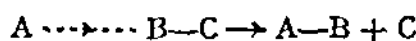
It should be noted that the rates of most reactions which can take place in solution but not in the gas phase are very much affected by the nature of the solvent (see also §5).

§3b. Transition state theory of reactions. London (1928, 1929) suggested a means of calculating energies of activation by methods of quantum mechanics. By making certain approximations, London developed an equation giving the variation with interatomic distances of the potential energy, E, of a system of three atoms. The equation is of the type $E = \int f \, dr$, and on solving for various distances r, a *potential energy surface* can be obtained which gives the variation of the potential energy for all possible interatomic distances. Thus the reaction between the three atoms A, B and C, *e.g.*,



$$A + BC \longrightarrow AB + C$$

must follow a path on the potential energy surface. According to London, the best approach of A to BC, *i.e.*, the path requiring the minimum energy, is for A to approach along the bonding line of BC and on the side remote from C, *i.e., a three-centre reaction has an end-on approach.*



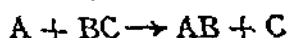
$$A \cdots\!\rightarrow\!\cdots B{-}C \longrightarrow A{-}B + C$$

Furthermore, BC repels A, and this repulsion prevents the close approach of A and so tends to prevent BC from reacting with A. In this three-centre reaction, the value of E depends on four factors: (i) the strength of the B—C bond; (ii) the repulsion between A and BC; (iii) the repulsion between AB and C; (iv) the strength

of the A—B bond. It therefore follows that charges on A and C, if any, and the dipoles of BC and AB, if any, will all affect the value of E, and consequently the approach of A to BC. It is important to note that in a three-centre end-on reaction, the configuration of B in the product is inverted.

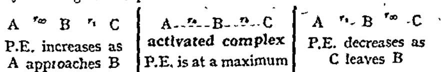
When we consider the mechanism of activation of this three-centre reaction, we can imagine that there are two extreme cases possible: (i) A is forced up against the repulsion of BC until it is close enough to compete with B on equal terms with C, which is finally expelled; (ii) BC acquires so much energy that the bond B—C is broken, and then A and B combine without any opposition.

Polanyi *et al.* (1931, 1935, 1938) amplified London's ideas into the *transition state theory*. These authors showed by mathematical treatment that it is usually more economical, *i.e.*, a lower value of E is necessary, for the reaction to proceed by a compromise between the two extremes mentioned above. A approaches BC along the bonding line of BC remote from C, and is forced against the repulsion of BC, and at the same time bond BC stretches until A and C can compete on equal terms for B. Thus a point is reached when the distances AB and BC are such that the forces between each pair are the same. This condition is the *transition state* (*activated complex*); in this state neither molecule AB nor BC exists independently. The system can now proceed in either direction to form A and BC, or AB and C. Thus the reaction



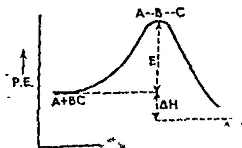
$$A + BC \rightarrow AB + C$$

may be imagined to proceed as follows:



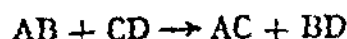
A $^{r_\infty}$ B ^{r_1} C	A--^{r_2}--B--^{r_2}--C	A ^{r_1} B $^{r_\infty}$ C
P.E. increases as A approaches B	activated complex P.E. is at a maximum	P.E. decreases as C leaves B

This sequence of events may be represented graphically (Fig. 1),



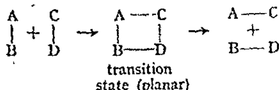
where E is the energy of activation and ΔH is the heat of reaction at constant pressure.

Four atom systems. Four-centre reactions, *i.e.*, reactions of the type



$$AB + CD \rightarrow AC + BD$$

have not been studied so extensively as the three-centre type. Eyring has suggested, from theoretical considerations, that the best approach, *i.e.*, the one which requires the minimum amount of energy, is a *side-approach* (*broadside collision*). AB and CD are in the same plane, with A near to C and B near to D; all four are involved in the transition state.

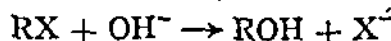


In reactions of this type, no inversion of C or D occurs. On the other hand, it may be possible for a four-centre reaction to proceed by an alternative path, *e.g.*, one molecule ionises first and then an ion reacts with the other molecule; in this case, the reaction is effectively a three-centre reaction, and so will lead to inversion.

In the transition state theory, it is assumed that the reaction rate is given by the rate at which the reacting molecules pass through the transition state. In this theory, the energy of activation is defined as the minimum amount of energy the reactant molecules must acquire in order to form an activated complex. If only the reactants are solvated (in solution), then the energy of activation is increased, and consequently the reaction rate is decreased. If only the activated complex is solvated, then E is decreased, and consequently the reaction rate is increased. If both the reactants and the activated complex are solvated, then there is often very little change in E, and consequently the influence of the solvent on the reaction rate is very small. In general, polar solvents favour reactions in which the activated complex is more polar than the reactants (see also §5).

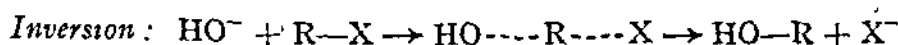
§4. Mechanism of the Walden inversion. We have, so far, considered the general theories of reaction mechanisms. We shall now apply these theories to the problem of the Walden inversion. Many theories have been suggested for the mechanism of the Walden inversion, but we shall discuss only those which are now generally accepted. The common feature of these theories is the intermediate formation of the transition state. If the replacing

group attacks from the position *furthest* from the group being replaced, then inversion will result. If the attack occurs on the *same side*, then no inversion will result. The essential problem, then, is to consider the forces which determine the direction of attack. Let us consider the reaction

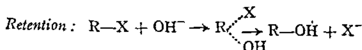


$$RX + OH^- \rightarrow ROH + X^-$$

If the attack occurs on the side remote from X, then inversion will result, if on the same side as X, then no inversion results. Thus:



Inversion: $HO^- + R{-}X \rightarrow HO{\cdots}R{\cdots}X \rightarrow HO{-}R + X^-$



Retention: $R{-}X + OH^- \rightarrow R\begin{smallmatrix}X\\OH\end{smallmatrix} \rightarrow R{-}OH + X^-$

In these reactions, R is an asymmetric radical in which the asymmetric carbon atom is joined to X (*cf.* §1). Polanyi *et al.* (1932) suggested that the polar bond C—X causes the negative ion (such as OH^-) to approach the molecule RX from the side remote from X (Fig. 2).

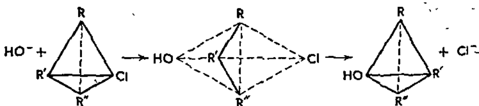
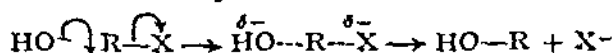


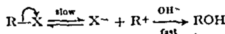
Fig. 3.2.

On the other hand, Hughes and Ingold (1937) have suggested from quantum mechanical arguments that, independently of the above electrostatic repulsions, the minimum energy of activation results when the attacking ion approaches from the direction that would lead to inversion Furthermore, these authors believe that the quantum mechanical forces are more powerful than the electrostatic ones (*cf.* §2 I). The Hughes–Ingold theory assumes that nucleophilic substitution reactions may take place by either of the following mechanisms: (i) An S_N2 mechanism involving the formation of an activated complex, *e.g.*,



$$HO^- + R{-}X \rightarrow \overset{\delta-}{HO}{\cdots}R{\cdots}\overset{\delta-}{X} \rightarrow HO{-}R + X^-$$

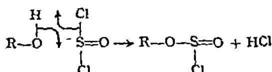
(ii) An S_N1 mechanism involving the formation of a carbonium ion, *e.g.*,



$$R{-}X \underset{}{\overset{\text{slow}}{\rightleftharpoons}} X^- + R^+ \xrightarrow[\text{fast}]{OH^-} ROH$$

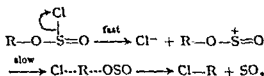
Hughes *et al.* (1935) studied (*a*) the interchange reaction of (±)-2-iodo-octane with radioactive iodine (as sodium iodide) in acetone solution, and (*b*) the racemisation of (+)-2-iodo-octane by ordinary sodium iodide under the same conditions. The reactions were shown to take place by an S_N2 mechanism, and the rates of exchange (of iodine) and racemisation were shown to be the same, *i.e.*, the halide-halide* displacement is always accompanied by inversion. Thus this experiment leads to the *assumption* that an S_N2 reaction always gives inversion. This is fully supported by other experimental work.

The reaction between alcohols and thionyl chloride has been studied extensively. According to Hughes and Ingold (1937), the first step in the reaction is the formation of a chlorosulphinate; no inversion occurs at this stage (which is a four-centre reaction)

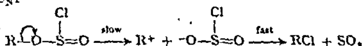


This intermediate chlorosulphinate may then form the alkyl chloride by one of the three possible mechanisms, S_N2, S_N1, or S_Ni; an S_Ni reaction is a nucleophilic reaction that takes place by an *intramolecular* mechanism.

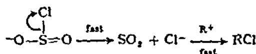
(i) S_N2



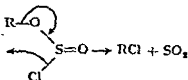
(ii) S_N1



The second stage of this reaction may possibly occur as follows:



(iii) S_Ni



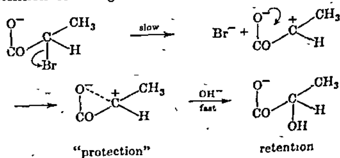
Hughes and Ingold based their arguments on the assumption that the alkyl chloride is formed by the decomposition of an intermediate chlorosulphinate, a compound which has been isolated in many cases, *e.g.*, from the reaction between thionyl chloride and methanol, ethanol, propanol, etc.

Now let us consider the steric course of each of the three possible mechanisms. If the reaction proceeds by the S_N2 mechanism, the reaction will be end-on, and so will be accompanied by inversion; an example of this type of reaction is that between thionyl chloride and octan-2-ol.

If the reaction proceeds by the S_N1 mechanism, *i.e.*, by pre-ionisation, then inversion and racemisation (retention) will occur, the amount of each depending on various factors. The carbonium ion is *flat* [*cf.* §8(iii). II], and therefore attachment of the chlorine ion can take place equally well on either side, *i.e.*, equal amounts of the (+)- and (−)-forms are produced; this is racemisation. During the actual ionisation, however, the retiring negative group will "protect" the carbonium ion from attack on that side, *i.e.*, there is a shielding effect, and this encourages an end-on attack *on the other side*, thereby leading to inversion. The amount of inversion and racemisation (retention) depends on the nature of the alkyl group, R, and on the experimental conditions, *e.g.*, the reaction between thionyl chloride and methylphenylmethanol, $CH_3 \cdot CHOH \cdot C_6H_5$, leads to inversion and racemisation (Hughes and Ingold). In this example, pre-ionisation is believed to be due to the powerful electron-releasing properties of the phenyl group.

Finally, if the reaction proceeds by intramolecular nucleophilic substitution (S_Ni mechanism), there is retention of configuration; the reaction proceeds effectively as a four-centre one.

In certain compounds involving attack at an asymmetric carbon atom, owing to the nature of one of the groups attached to the asymmetric carbon atom, an S_N1 reaction occurs with almost complete retention of configuration. This retention is due to "pro-



tection", and the most important "configuration-protecting" group is the α-carboxylate ion group. An example which illustrates this phenomenon is the hydrolysis of α-bromopropionic acid to α-hydroxypropionic acid in alkaline solution.

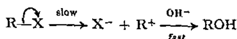
The negatively charged oxygen atom of the carboxylate ion forms a "weak electrostatic bond" with the positively charged carbon atom on the side remote from that where the bromine ion is expelled. Thus this remote side is protected from attack by the hydroxyl ion, and consequently the hydroxyl ion is forced to attack on the same side as that of the receding bromine ion, thereby leading mainly to retention of configuration (Hughes and Ingold, 1937).

Summary. The Walden inversion affords a means of studying the mechanism of substitution and replacement reactions. If complete inversion occurs, then the reaction takes place by an S_N2 mechanism; if racemisation and inversion (the latter generally being in excess), then the mechanism is S_N1. When the amount of racemisation is small and there is a predominant amount of retention, the mechanism is also S_N1, but in this case a "configuration-protecting" group is present. When the reaction takes place with complete retention, the mechanism is S_Ni; complete retention, however, also occurs with four-centre reactions. These rules also offer a means of correlating configurations.

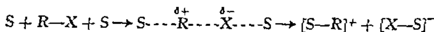
The Walden inversion has also afforded a means of studying the mechanisms of various rearrangements. Thus the experiments of Kenyon *et al.* (1939, 1941, 1945) on the mechanisms of the Hofmann, Curtius, Lossen and Beckmann rearrangements (see Vol. I), have shown that these rearrangements occur with retention of configuration when the migrating group contains an asymmetric carbon atom; thus these rearrangements can be depicted by an S_Ni mechanism (see also §2h. VI).

§5. Effect of solvents on reaction velocity and mechanism of reaction. Since the rate-determining step in an S_N1 reaction is the ionisation into the carbonium ion, any factor that assists this ionisation will therefore tend to help S_N1 reactions to take place more readily. Solvents with high dipole moments are usually good ionising media, and in general it is found that the more polar the solvent, the greater the velocity of an S_N1 reaction. However, in addition to the ionisation effect, there is also the problem of solvation to be considered. Ions and polar molecules, when dissolved in polar solvents, tend to become solvated. For a given solvent, the solvation tends to increase with the increasing magnitude of the charge on the solute molecules or ions; and for a given solute, the

solvation tends to increase with the increasing dipole moment of the solvent. Thus, in the S_N1 reaction:

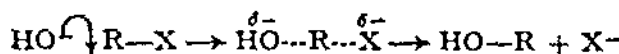


the first step, in a highly polar solvent, may be written:



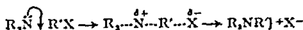
where S represents a number of solvent molecules. Both the high dipole moment of the solvent and the process of solvation will facilitate the ionisation of RX. On the other hand, the solvated ion $[S—R]^+$ would probably not be so active as the non-solvated ion R^+, and so the second step is slowed down. This retardation is of much less magnitude than the acceleration of the first step, and so the overall reaction proceeds at a faster rate in highly polar solvents. It should be noted that the above reaction is an S_N1 reaction only so long as the first step is slower than the second. Should the second become slower than the first, then the mechanism will be S_N2 (see also below).

The rates of S_N2 reactions are also affected by the polarity of the solvent. In addition to the assumption that the greater the polarity of the solute molecules, the greater will be the solvation, there is also the assumption made (and is borne out in practice) that for a given magnitude of charge, solvation decreases as the charge is spread over a larger volume. Thus, in the S_N2 reaction



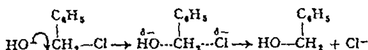
a solvent with a high dipole moment will solvate both the reactant ion and the transition state, but more so the former than the latter, since in the latter the charge is more dispersed than in the former. Thus solvation tends to stabilise the reactants more than the transition state, *i.e.*, the energy of activation is increased, and consequently the reaction is retarded.

In the Menschutkin reaction:



the charge on the transition state is greater than that on the reactants. Thus the transition state will be more solvated than the reactants, thereby lowering the energy of activation, and thus increasing the rate of the reaction. Furthermore, the greater the polarity of the solvent, the more readily solvation should occur. These predictions have been observed experimentally, *the rate of the reaction increasing with increasing polarity of the solvent.*

The polarity of the solvent may change the mechanism of a reaction, *e.g.*, Olivier (1934) showed that the hydrolysis of benzyl chloride in 50 per cent. aqueous acetone is largely by the S_N2 mechanism.



In water, however, the mechanism was changed to mainly S_N1. The dipole moment of water is greater than that of aqueous acetone, and consequently ionisation of the benzyl chloride is facilitated.

In general, the nature of the reactant molecules and solvent determines which mechanism will take place (*cf.* the work of the earlier investigators, §2).

§6. Nature of the transition state. In the reaction



we have seen that in the transition state X and Y are collinear and on opposite sides of the attacked carbon atom. In the transition state, C, *a*, *b* and *d* are coplanar, with the line joining X and Y perpendicular to this plane. In the original molecule C*abd*X, the four groups are arranged tetrahedrally. Hence to achieve a planar configuration of C*abd* in the transition state, the carbon atom changes from tetrahedral to trigonal hybridisation (see Vol. I, Ch. II), the remaining p_z orbital being used (by means of its two lobes) to hold the groups X and Y by "half-bonds" (Fig. 3).

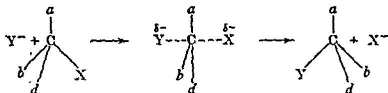
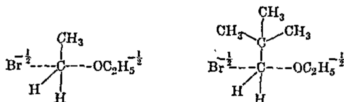


Fig. 3.3

When X is ejected, the carbon atom returns to its state of tetrahedral hybridisation.

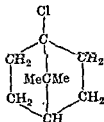
§6a. Steric factors and the transition state. We have already seen (§4. II) that steric strain, produced by steric repulsion, will decrease the stability of a molecule. This steric effect may also operate in the transition state, and the compression energy in the transition state may be either greater or smaller than that in the initial state and consequently the steric effects may hinder or

facilitate the reaction. In the S_N2 mechanism, the transition state consists of five atoms or groups bonded or partly bonded to the carbon atom at which reaction is occurring. If these groups are bulky then steric repulsions between them prevent the ready formation of the transition state, *i.e.*, the compression energy in the transition state is greater than that in the initial state, and so the reaction is hindered sterically. A classical example is the case of the *neo*pentyl radical, *e.g.*, the S_N2 reaction involving *neo*pentyl bromide and sodium ethoxide takes place with far greater difficulty than the analogous reaction involving ethyl bromide. The transition states of these two reactions may be written as follows:

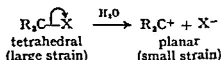


In the " normal " transition state, the entering and displaced groups are collinear. Such a state of affairs is possible with ethyl bromide, but is not possible with *neo*pentyl bromide because of the large bulk of the *tert.*-butyl group. In the latter case it is believed that the Br—C—OC_2H_5 bonds are no longer collinear but " bent away " from the *tert.*-butyl group. Such a " bent " transition state has a large compression energy and so is far more difficult to form than a symmetrical (" normal ") transition state.

Another interesting example of steric effects on the transition state is that of 1-chloroapocamphane. This compound does not react with reagents that normally react with alkyl halides (Bartlett *et al.*, 1939). The hydrolysis of *tert.*-butyl chloride always takes place by the S_N1 mechanism. 1-Chloroapocamphane is a tertiary chloride, but it does not undergo ionisation and so an S_N1 mechanism is not possible. The reason for this failure to ionise is believed to be as follows. Carbonium ions are flat, and if any structural feature prevents a planar configuration, then no carbonium ion will be produced. Removal of the chloride ion from 1-chloroapocamphane would produce a positively charged carbon atom which cannot become planar because of the steric requirements of the ring structure. On the other hand, since the rear of the carbon atom of the C—Cl group is protected by the bridge, an S_N2 mechanism is not possible since inversion cannot occur.



In general, steric hindrance is of far less importance in S_N1 mechanisms. This is due to the fact that in the S_N1 mechanism there are not more than four groups attached to the carbon atom at which reaction is occurring (but see solvation, §5). If, however, the molecule (undergoing the S_N1 reaction) contains large groups, then the first step of ionisation may relieve the steric strain in that molecule and so assist the formation of the carbonium ion, *i.e.*, the reaction may be *accelerated sterically*. Thus, in the solvolysis of tertiary halides we have:



Brown *et al.* (1949) have shown that as the R groups increase in size, the rate of solvolysis increases. However, the larger the R groups, the more slowly will the carbonium ion be expected to react with the solvent, and thus a factor is introduced which opposes steric acceleration. As we have seen, a carbonium ion can also undergo an elimination reaction to form an olefin, and Brown *et al.* (1950) have shown that this elimination process also increases as the R groups become larger.

ASYMMETRIC SYNTHESIS

§7. Partial asymmetric synthesis. Partial asymmetric synthesis may be defined as a method for preparing optically active compounds from symmetrical compounds by the intermediate use of optically active compounds, but without the necessity of resolution (Marckwald, 1904). In ordinary laboratory syntheses, a symmetrical compound always produces the racemic modification (§7a. II).

The first asymmetric synthesis was carried out by Marckwald (1904), who prepared an active (—)-valeric acid (lævorotatory to the extent of about 10 per cent. of the pure compound) by heating the half-brucine salt of ethylmethylmalonic acid at 170°.

I and II are diastereoisomers; so are III and IV. V and VI are enantiomorphs, and since the mixture is optically active, they must be present in unequal amounts. Marckwald believed this was due to the different rates of decomposition of diastereoisomers I and II, but according to Eisenlohr and Meier (1938), the half-brucine salts I and II are not present in *equal* amounts in the solid form (as thought by Marckwald). These authors suggested that as the less soluble diastereoisomer crystallised out (during evaporation of the

$$\mathrm{CH_3(C_2H_5)C(CO_2H)_2} \xrightarrow{(-)\text{-brucine}}$$

$$\underset{\text{I}}{\mathrm{CH_3(C_2H_5)C(CO_2H)CO_2H[(-)\text{-brucine}]}} + \underset{\text{II}}{\mathrm{CH_3(C_2H_5)C(CO_2H)CO_2H[(-)\text{-brucine}]}}$$

$$\xrightarrow{170^\circ} \underset{\text{III}}{\mathrm{CH_3(C_2H_5)C(H)CO_2H[(-)\text{-brucine}]}} + \underset{\text{IV}}{\mathrm{CH_3(C_2H_5)C(H)CO_2H[(-)\text{-brucine}]}}$$

$$\xrightarrow{\text{HCl}} \underset{\text{V}}{\mathrm{CH_3(C_2H_5)C(H)CO_2H}} + \underset{\text{VI}}{\mathrm{CH_3(C_2H_5)C(H)CO_2H}}$$

solution), some of the more soluble diastereoisomer spontaneously changed into the less soluble diastereoisomer to restore the equilibrium between the two; thus the final result was a mixture of the half-brucine salt containing a larger proportion of the less soluble diastereoisomer. If this be the explanation, then we are dealing with an example of asymmetric transformation and not of asymmetric synthesis (see §10. II). Further work, however, has shown that Marckwald had indeed carried out an asymmetric synthesis. Kenyon and Ross (1951) decarboxylated optically active ethyl hydrogen ethylmethylmalonate, VII, and obtained an optically inactive product, ethyl ($\pm$)-α-methylbutyrate, VIII.

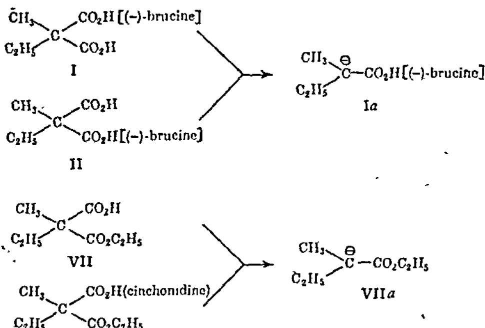
$$\underset{\substack{\text{VII}\\ \text{active}}}{\mathrm{CH_3(C_2H_5)C(CO_2H)CO_2C_2H_5}} \longrightarrow CO_2 + \underset{\substack{\text{VIII}\\ \text{inactive}}}{\mathrm{CH_3(C_2H_5)C(H)CO_2C_2H_5}}$$

These authors (1952) then decarboxylated the cinchonidine salt of VII, and still obtained the optically inactive product VIII.

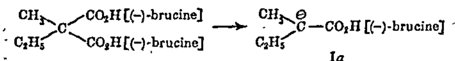
$$\mathrm{CH_3(C_2H_5)C(CO_2C_2H_5)CO_2H(cinchonidine)} \longrightarrow \underset{\substack{\text{VIII}\\ \text{inactive}}}{\mathrm{CH_3(C_2H_5)C(H)CO_2C_2H_5}} + CO_2 + \text{cinchonidine}$$

Kenyon and Ross suggest the following explanation to account for their own experiments and for those of Marckwald. Decarboxyla-

tion of diastereoisomers I and II takes place *via* the formation of the same carbanion I*a*, and decarboxylation of VII and its cinchonidine salt *via* VII*a*



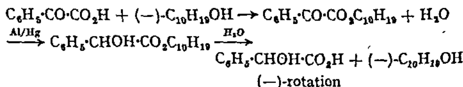
Combination of carbanion I*a* with a proton will produce diastereoisomers III and IV in different amounts, since, in general, diastereoisomers are formed at different rates (§7*b*. II). On the other hand, carbanion VII*a* will give equimolecular amounts of the enantiomorphs of VIII. If the formation of optically active α-methylbutyric acid (V and VI) were due to different rates of decarboxylation of III and IV (Marckwald's explanation) or to partial asymmetric transformation during crystallisation (Eisenlohr and Meier's explanation), then these effects are nullified if Kenyon's explanation is correct, since the intermediate carbanion is the *same* for both diastereoisomers. Thus, if the asymmetric transformation theory were correct, then decarboxylation of the *dibrucine* salt of ethylmethylmalonic acid to α-methylbutyric acid should give an optically inactive product, since only one type of crystal is now possible (asymmetric transformation is now impossible).



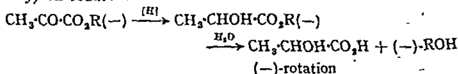
On the other hand, if the carbanion I*a* is an intermediate in this

decomposition, it is still possible to obtain an optically active product. Kenyon and Ross did, in fact, obtain a lævorotatory product.

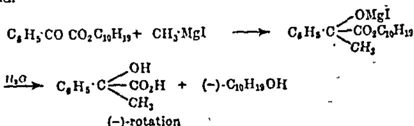
McKenzie (1904) carried out a number of partial asymmetric syntheses by reduction of the keto group in various keto-esters in which the ester group contained an asymmetric group, *e.g.*, benzoylformic acid was esterified with (—)-menthol, the ester reduced with aluminium amalgam, and the resulting product saponified; the mandelic acid so obtained was slightly lævorotatory.



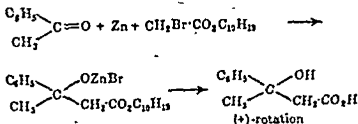
Similarly, the pyruvates of (—)-menthol, (—)-pentyl alcohol and (—)-borneol gave an optically active lactic acid (slightly lævorotatory) on reduction.



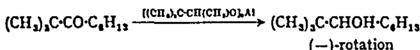
McKenzie (1904) also obtained similar results with Grignard reagents, *e.g.*, the (—)-menthyl ester of benzoylformic acid and methylmagnesium iodide gave a slightly lævorotatory atrolactic acid.



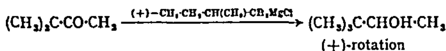
Turner *et al.* (1949) carried out a Reformatsky reaction (see Vol. I) using acetophenone, (—)-menthyl bromoacetate and zinc, and obtained a dextrorotatory β-hydroxy-β-phenylbutyric acid.



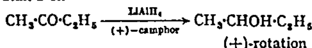
Jackman *et al.* (1950) reduced *tert.*-butyl *n*-hexyl ketone with aluminium (+)-1:2:2-trimethylpropoxide at 200°, and obtained a slightly lævorotatory alcohol.



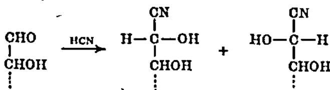
Another example of asymmetric synthesis involving the use of a Grignard reagent is the *reduction* of 3·3-dimethylbutan-2-one into a dextrorotatory 3:3-dimethylbutan-2-ol by means of (+)-2-methylbutylmagnesium chloride (Mosher *et al.*, 1950; see also Vol. I for abnormal Grignard reactions).



In all the examples given on asymmetric synthesis, the optically active agent used as an intermediate has been combined with one or other of the reacting molecules. Bothner-By (1951), however, has carried out an asymmetric synthesis where the optically active agent is present but not combined with either of the reacting molecules. This author reduced butanone with lithium aluminium hydride in the presence of (+)-camphor, and thereby obtained (+)-*iso*borneol (from the camphor) and a small amount of a dextrorotatory butan-2-ol.

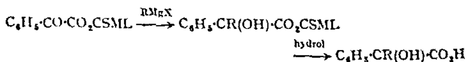


It has already been pointed out that a molecule containing one asymmetric carbon atom gives rise to a pair of diastereoisomers in *unequal* amounts when a second asymmetric carbon atom is introduced into the molecule (§7b. II). In general, if a new asymmetric centre is introduced into a molecule which is already asymmetric, the asymmetric part of the molecule influences the configuration formed from the symmetrical part of the molecule, the two diastereoisomers being formed in unequal amounts, *e.g.*, the Kiliani reaction (see also Vol. I).



Prelog *et al.* (1953) have studied, by means of conformational

analysis, the steric course of the addition of Grignard reagents to benzoylformic (phenylglyoxylic) esters of asymmetric alcohols. If the letters S, M and L refer respectively to small, medium and large groups attached to the carbinol carbon atom of the asymmetric alcohol, then the general reaction may be written:



$$C_6H_5{\cdot}CO{\cdot}CO_2CSML \xrightarrow{RMgX} C_6H_5{\cdot}CR(OH){\cdot}CO_2CSML \xrightarrow{hydrol} C_6H_5{\cdot}CR(OH){\cdot}CO_2H$$

Prelog *et al* found that the configuration of the asymmetric carbon atom in the stereoisomer that predominated in this reaction could be correlated with that of the carbinol carbon of the alcohol. The basis of this correlation was the assumption that the Grignard reagent attacks the carbon atom (of the ketone group) preferentially from the less hindered side. This necessitates a consideration of the possible conformations of the ester molecule. The authors considered that the most stable conformation of the ester was the one in which the two carbonyl groups are planar and *trans* to each other, with the smallest group lying in this plane and the other two groups skew. Furthermore, with the groups on the carbinol atom of the alcohol arranged in the staggered conformation with respect to the rest of the molecule, then IX and X will be the conformations of the esters with the enantiomorphous alcohol residues IX*a* and X*a* respectively (thick lines represent groups in front of the plane, broken lines groups, behind, and ordinary lines groups in the plane). Thus, with L behind, methylmagnesium halide attacks preferentially from the front (IX), and with L in front, the attack is from behind (X). The α-hydroxyacid obtained from IX is IX *b*, and that from X is X *b*. IX *b* and X *b* are enantiomorphs and hence the configuration of the new asymmetric centre is related to that of the adjacent asymmetric centre in the original molecule. Thus for the same keto-acid and the same Grignard reagent, and using different optically active alcohols belonging to the same configurational series, the product should contain excess of α-hydroxyacids with the same sign of rotation. This has been shown to be so in practice, *e.g.*, (−)-menthol and (−)-borneol are both configurationally related to L(−)-glyceraldehyde, and both lead to a predominance of the (−)-hydroxyacid. On the other hand, if the keto-acid is pyruvic acid and the Grignard reagent phenylmagnesium bromide, the (+)-hydroxyacid should predominate in the product (this method of preparation produces an interchange of the positions of the phenyl and methyl groups, thereby leading to the formation of the enantiomorph). This can be seen from the following equations, starting

M
HO—C—S
L
IX *a*

Ph—C(=O)—C(=O)—O—C(L)(M)(S)
MeMgX
IX

Ph—C(OH)(Me)—C(=O)—O—C(L)(M)(S)

CO_2H
Me—C—OH
Ph
IX *b*

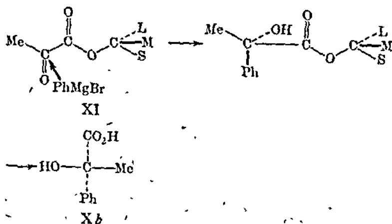
M
S—C—OH
L
X *a*

Ph—C(=O)—C(=O)—O—C(M)(L)(S)
MeMgX
X

Ph—C(Me)(OH)—C(=O)—O—C(M)(L)(S)

CO_2H
HO—C—Me
Ph
X *b*

with the pyruvic ester XI in which the configuration of the alcohol is IX *a*, the product would be X *b*.

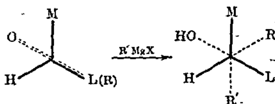


These results have been obtained in practice. Thus, when the configuration of the active alcohol is known, it is possible to deduce the configuration of the α-hydroxyacid obtained in excess. This method has been used to determine the configuration of hydroxyl groups in steroids.

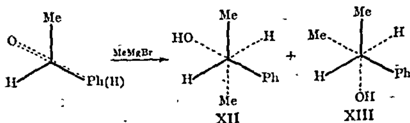
Cram *et al.* (1952) have also dealt with asymmetric syntheses in which the molecule contains an asymmetric centre that belongs to the molecule, *i.e.*, remains in the molecule (*cf.* the Kiliani reaction mentioned above). As a result of their work, these authors have formulated the rule of "steric control of asymmetric induction". This is: "In non-catalytic reactions of the type shown, that diastereoisomer will predominate which would be formed by the approach of the entering group from the *least hindered side* of the double bond when the rotational conformation of the C—C bond is such that the double bond is flanked by the two least bulky groups attached to the adjacent asymmetric centre." Thus:



or, using the projection formulae:



An example of this type of reaction is the reaction between phenylpropionaldehyde (M = Me, L = Ph) and methylmagnesium bromide (R′ = Me); two products can be formed, *viz.*, XII the (*erythro*-compound) and XIII (the *threo*-compound):



According to the above rule, XII should predominate; this has been found to be so in practice.

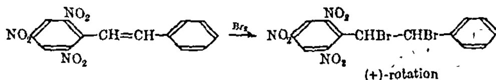
The influence of enzymes on the steric course of reactions has also been investigated, *e.g.*, Rosenthaler (1908) found that emulsin converted benzaldehyde and hydrogen cyanide into dextrorotatory mandelonitrile which was almost optically pure. It has been found that in most enzymic reactions the product is almost 100 per cent. of one or other enantiomorph. Enzymes are proteins and optically active (see also §12. XIII), but since they are so "one-sided" in their action, it appears likely that the mechanism of the reactions in which they are involved differs from that of partial asymmetric syntheses where enzymes are not used. It has been suggested that enzymes are the cause of the formation of optically active compounds in plants. Although this is largely true, the real problem is: How were the optically active enzymes themselves produced? Ferreira's work [§10(viii). II], however, shows that optically active compounds may possibly be produced in living matter by activation of a racemic modification. This theory appears to be superior to that of the formation of optically active compounds by the action of naturally polarised light (see following section).

§8. Absolute asymmetric synthesis. Cotton (1896) found that dextro- and lævocircularly polarised light was unequally absorbed by enantiomorphs, provided the light has a wavelength in the neighbourhood of the characteristic absorption bands of the compound. This phenomenon is known as the **Cotton effect** or **circular dichroism** (*cf.* §2. II).

It has been suggested that circularly polarised light produced the first natural active compounds, and to support this theory, racemic modifications have been irradiated with circularly polarised light and attempts made to isolate one enantiomorph. There was very little success in this direction until W. Kuhn and Braun (1929) claimed to have obtained a small rotation in the case of ethyl α-bromopropionate. The racemic modification of this compound was irradiated with right- and left-circularly polarised light (of wavelength 2800 A), and the product was found to have a rotation of + or − 0·05°, respectively. Thus we have the possibility of preparing optically active products from inactive substances *without* the intermediate use of optically active reagents (*cf.* Ferreira's work). This type of synthesis is known as an **absolute asymmetric synthesis**; it is also known as an **absolute asymmetric decomposition.** The term asymmetric decomposition is also applied to reactions such as the formation of the (+)- and (−)-forms of αγ-di-1-naphthyl-αγ-diphenylallene (see §6. V) by the action of (+)- and (−)- camphorsulphonic acid on the symmetrical alcohol.

From 1930 onward, more conclusive evidence for absolute asymmetric syntheses has been obtained, *e.g.*, W. Kuhn and Knopf (1930) irradiated (±)-α-azidopropionic dimethylamide, $CH_3 \cdot CHN_3 \cdot CO \cdot N(CH_3)_2$, with right-circularly polarised light and obtained an undecomposed product with a rotation of +0·78°; with left-circularly polarised light, the undecomposed product had a rotation of −1·04°. Thus the (−)- or (+)-form is decomposed (photochemically) by right- or left-circularly polarised light, respectively. Similarly, Mitchell (1930) irradiated humulene nitrosite with right- and left-circularly polarised red light, and obtained slightly optically active products

Davis and Heggie (1935) found that the addition of bromine to 2:4·6-trinitrostilbene in a beam of right-circularly polarised light gave a dextrorotatory product.



Small (+)-rotations were also observed when a mixture of ethyl fumarate and anhydrous hydrogen peroxide in ethereal solution was irradiated with right-circularly polarised light (Davis *et al.*, 1945).

Reading References

Hinshelwood, *The Kinetics of Chemical Change*, Oxford Press (1940, 4th ed.).
Moelwyn-Hughes, *The Kinetics of Reactions in Solutions*, Oxford Press (1947, 2nd ed.).
Glasstone, Laidler and Eyring, *The Theory of Rate Processes*, McGraw-Hill (1941).
Frost and Pearson, *Kinetics and Mechanism*, Wiley (1953).
Friess and Weissberger (Ed.), *Technique of Organic Chemistry*, Interscience Publishers. Vol 8 (1953). Investigation of Rates and Mechanisms of Reactions.
Ingold, *Structure and Mechanism in Organic Chemistry*, Bell and Sons (1953). Ch. 2. Interactions between Molecules Ch 7. Nucleophilic Aliphatic Substitution.
Wheland, *Advanced Organic Chemistry*, Wiley (1949). Ch. 7 (p. 269). The Walden Inversion.
Hammett, *Physical Organic Chemistry*, McGraw-Hill (1940). Ch. 6. Stereochemistry of the Displacement Reaction.
Dewar, *The Electronic Theory of Organic Chemistry*, Oxford Press (1949). Ch. V. The Replacement Reaction.
Remick, *Electronic Interpretations of Organic Chemistry*, Wiley (1945). Chh. 7, 9, 10 and 11.

Waters, *Physical Aspects of Organic Chemistry*, Routledge and Kegan Paul (1950, 4th ed.). Ch 12. Substitution Reactions of Organic Halogen Compounds.

Ann. Reports (Chem. Soc.), 1953, **50**, 134 Unimolecular Nucleophilic Displacements and Related Reactions

Gold, Orbital Hybridisations and Some Other Considerations Concerning the Transition State of Bimolecular Substitution Reactions, *J.C.S.*, **1951**, 1430.

Ritchie, *Asymmetric Synthesis and Asymmetric Induction*, St Andrews University Press (1933).

Ritchie, Recent Views on Asymmetric Synthesis and Related Processes, *Advances in Enzymology*, Interscience Publishers, 1947, **7**, 65

Cram and Elhafez, The Rule of "Steric Control of Asymmetric Induction" in the Synthesis of Acyclic Systems, *J Amer Chem. Soc.*, 1952, **74**, 5828

Kenyon and Ross, A New Mechanism for the Marckwald Asymmetric Synthesis, *J.C.S.*, **1952**, 2307.

Klyne (Ed.), *Progress in Stereochemistry*, Butterworth (1954) Ch. 3. Stereochemical Factors in Reaction Mechanisms and Kinetics. Vol. II (1958). Chh. 2, 3.

Streitwieser, Solvolytic Displacement Reactions at Saturated Carbon Atoms, *Chem. Reviews*, 1956, **56**, 571.

From 1930 onward, more conclusive evidence for absolute asymmetric syntheses has been obtained, *e.g.*, W. Kuhn and Knopf (1930) irradiated (±)-α-azidopropionic dimethylamide, $CH_3·CHN_3·CO·N(CH_3)_2$, with right-circularly polarised light and obtained an undecomposed product with a rotation of +0·78°; with left-circularly polarised light, the undecomposed product had a rotation of −1·04°. Thus the (−)- or (+)-form is decomposed (photochemically) by right- or left-circularly polarised light, respectively. Similarly, Mitchell (1930) irradiated humulene nitrosite with right- and left-circularly polarised red light, and obtained slightly optically active products.

Davis and Heggie (1935) found that the addition of bromine to 2:4 6-trinitrostilbene in a beam of right-circularly polarised light gave a dextrorotatory product.

$$(NO_2)_3C_6H_2{-}CH{=}CH{-}C_6H_5 \xrightarrow{Br_2} (NO_2)_3C_6H_2{-}CHBr{-}CHBr{-}C_6H_5$$

(+)-rotation

Small (+)-rotations were also observed when a mixture of ethyl fumarate and anhydrous hydrogen peroxide in ethereal solution was irradiated with right-circularly polarised light (Davis *et al.*, 1945).

Reading References

Hinshelwood, *The Kinetics of Chemical Change*, Oxford Press (1940, 4th ed.).
Moelwyn-Hughes, *The Kinetics of Reactions in Solutions*, Oxford Press (1947, 2nd ed.).
Glasstone, Laidler and Eyring, *The Theory of Rate Processes*, McGraw-Hill (1941).
Frost and Pearson, *Kinetics and Mechanism*, Wiley (1953).
Friess and Weissberger (Ed.), *Technique of Organic Chemistry*, Interscience Publishers. Vol. 8 (1953). Investigation of Rates and Mechanisms of Reactions.
Ingold, *Structure and Mechanism in Organic Chemistry*, Bell and Sons (1953). Ch. 2. Interactions between Molecules Ch 7. Nucleophilic Aliphatic Substitution.
Wheland, *Advanced Organic Chemistry*, Wiley (1949). Ch 7 (p 269). The Walden Inversion.
Hammett, *Physical Organic Chemistry*, McGraw-Hill (1940). Ch 6. Stereochemistry of the Displacement Reaction.
Dewar, *The Electronic Theory of Organic Chemistry*, Oxford Press (1949). Ch V. The Replacement Reaction.
Remick, *Electronic Interpretations of Organic Chemistry*, Wiley (1945). Chh. 7, 9, 10 and 11.

Waters, *Physical Aspects of Organic Chemistry*, Routledge and Kegan Paul (1950, 4th ed.). Ch. 12. Substitution Reactions of Organic Halogen Compounds.

Ann. Reports (Chem. Soc.), 1953, **50**, 134. Unimolecular Nucleophilic Displacements and Related Reactions

Gold, Orbital Hybridisations and Some Other Considerations Concerning the Transition State of Bimolecular Substitution Reactions, *J.C.S.*, **1951**, 1430.

Ritchie, *Asymmetric Synthesis and Asymmetric Induction*, St Andrews University Press (1933).

Ritchie, Recent Views on Asymmetric Synthesis and Related Processes, *Advances in Enzymology*, Interscience Publishers, 1947, **7**, 65

Cram and Elhafez, The Rule of "Steric Control of Asymmetric Induction" in the Synthesis of Acyclic Systems, *J. Amer. Chem. Soc*, 1952, **74**, 5828

Kenyon and Ross, A New Mechanism for the Marckwald Asymmetric Synthesis, *J.C.S.*, **1952**, 2307.

Klyne (Ed.), *Progress in Stereochemistry*, Butterworth (1954). Ch 3 Stereochemical Factors in Reaction Mechanisms and Kinetics. Vol. II (1958). Chh. 2, 3.

Streitwieser, Solvolytic Displacement Reactions at Saturated Carbon Atoms, *Chem. Reviews*, 1956, **56**, 571.

the double bond is not to be expected. Furthermore, according to the above arrangement, the two hydrogen atoms and the two carboxyl groups are all in one plane, *i.e.*, the molecule is flat. Since a flat molecule is superimposable on its mirror image, maleic and fumaric acids are therefore not optically active (§2. II). As we shall see later, modern theory also postulates a planar structure for these two acids, but the reasons are very much different from those proposed by van't Hoff as described above (see §3).

The type of isomerism exhibited by maleic and fumaric acids is known as **geometrical isomerism** or ***cis-trans* isomerism.** One isomer is known as the *cis*-compound, and the other as the *trans*, the *cis*-compound being the one which (usually) has identical or similar atoms or groups, on the *same* side (see also §4). Thus molecule I is *cis*-butenedioic acid, and II is *trans*-butenedioic acid. As will be shown later (§5), I is maleic acid and II fumaric acid.

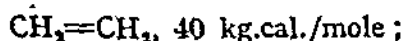
Geometrical isomerism is exhibited by a wide variety of compounds, and they may be classified into three groups:

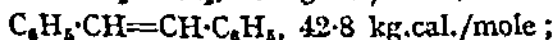
(i) Compounds containing a double bond: C=C, C=N, N=N.

(ii) Compounds containing a cyclic structure—homocyclic, heterocyclic and fused ring systems.

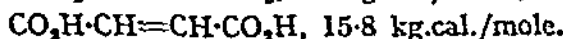
(iii) Compounds which may exhibit geometrical isomerism due to restricted rotation about a single bond (see §3. V for examples of this type).

§2. **Rotation about a double bond.** We have already seen that, theoretically, there is always some opposition to rotation about a *single* bond and that, in many cases, the opposition may be great enough to cause the molecule to assume some preferred conformation (§4a. II). When we consider the problem of rotation about a *double* bond, we find that there is always considerable opposition to the rotation. Let us first consider the simple case of ethylene; Fig. 2 (*a*) shows the energy changes in the molecule when one methylene group is rotated about the carbon-carbon double bond with the other methylene group at rest. Thus there are two *identical* favoured positions (one at 0° and the other at 180°), and the potential energy barrier is 40 kg.cal./mole. The examination of many olefinic compounds has shown that the potential energy barrier for the C=C bond varies with the nature of the groups attached to each carbon, *e.g.*,


$CH_2{=}CH_2$, 40 kg.cal./mole;


$C_6H_5{\cdot}CH{=}CH{\cdot}C_6H_5$, 42·8 kg.cal./mole;


$CH_3{\cdot}CH{=}CH{\cdot}CH_3$, 18 kg.cal./mole;

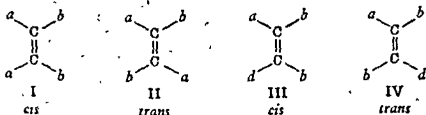

$CO_2H{\cdot}CH{=}CH{\cdot}CO_2H$, 15·8 kg.cal./mole.

Let us consider the case of maleic and fumaric acids in more detail.

configuration the two p_z orbitals are parallel. Distortion of the molecule from the planar configuration decreases the overlap of the π-electrons, thereby weakening the π-bond; and this distortion can only be effected by supplying energy to the molecule. It is therefore this tendency to produce **maximum overlap of the π-electrons in the π-bond** that gives rise to resistance of rotation about a "double" bond. For simplicity we shall still represent a "double" bond by the conventional method, *e.g.*, C=C, but it should always be borne in mind that *one* of these bonds is a σ-bond (sp^2 bond), and the *other* is a π-bond perpendicular to the σ-bond. It is these π-electrons (*mobile electrons*) which undergo the electromeric and resonance effects. They are held less firmly than the σ-electrons and are more exposed to external influences; it is these π-electrons which are responsible for the high reactivity of unsaturated compounds.

In compounds containing a triple bond, *e.g.*, acetylene, the two carbon atoms are in a state of *digonal* hybridisation; there are two σ-bonds (sp bonds) and two π-bonds (one p_y and one p_z orbital), both perpendicular to the σ-bonds which are collinear (see Vol. I, Ch. II).

§4. Nomenclature of geometrical isomers. When geometrical isomerism is due to the presence of *one* double bond in a molecule, it is easy to name the geometrical isomers if two groups are identical, *e.g.*, in molecules I and II, I is the *cis*-isomer, and II the *trans*; similarly III is *cis*, and IV is *trans*. When, however,



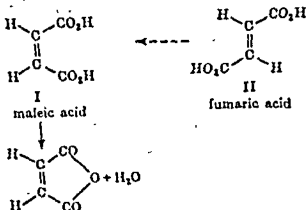
all four groups are different, nomenclature is more difficult. In this case it has been suggested that the prefixes *cis* and *trans* should indicate the disposition of the *first two* groups named, *e.g.*, the two stereoisomers of 1-bromo-1-chloro-2-iodoethylene, V and VI; V is *cis*-1-bromo-2-iodo-1-chloroethylene or *trans*-1-chloro-2-iodo-1-bromoethylene; VI is *cis*-1-chloro-2-iodo-1-bromoethylene or *trans*-1-bromo-2-iodo-1-chloroethylene. On the other hand, since this method of nomenclature usually deviates from the rule of naming groups in alphabetical order, it has been suggested that the groups corresponding to the prefix *cis* or *trans* should be italicised, thus V

The number of geometrical isomers is 2^n, where n is the number of double bonds; this formula applies only to molecules in which the ends are different. If the ends are identical, *e.g.*, CHa=CH—CH=CHa, then the number of stereoisomers is $2^{n-1} + 2^{p-1}$, where $p = n/2$ when n is even, and $p = \frac{n+1}{2}$ when n is odd (Kuhn *et al.*, 1928).

§5. Determination of the configuration of geometrical isomers. There is no general method for determining the configuration of geometrical isomers. In practice one uses a number of different methods, the method used depending on the nature of the compound in question. The following are methods which may be used mainly for compounds that owe their geometrical isomerism to the presence of a double bond, but several of the methods are special to geometrical isomers possessing a cyclic structure (see also §7).

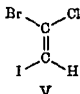
Method of cyclisation. ... first to suggest ... likely to occur ... molecule. This ... which *rings* are formed, but does not hold for elimination reactions in which a double (or triple) bond is produced [see, *e.g.*, (xi)].

(*a*) Of the two acids maleic and fumaric, only the former readily forms a cyclic anhydride when heated; the latter does not form an anhydride of its own, but when strongly heated, gives maleic anhydride. Thus I is maleic acid, and II is fumaric acid.

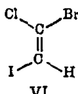


Cyclisation reactions must be performed carefully, since one may be converted into the other during the cyclising process, ... to unreliable results. In the above reaction, somewhat ...ditions have been used; hence there is the possibility ...ersion of the stereoisomers has occurred. Since ... readily, and fumaric acid only after prolonged

may be named *cis*-1-*bromo*-1-chloro-2-*iodo*ethylene and VI *trans*-1-*bromo*-1-chloro-2-*iodo*ethylene. This method, it must be admitted, would offer difficulties when the names are spoken.



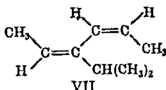
V



VI

Some pairs of geometrical isomers have trivial names, *e.g.*, maleic and fumaric acids, angelic and tiglic acids, etc. (*cf.* §1). Sometimes the prefix *iso* has been used to designate the *less* stable isomer, *e.g.*, crotonic acid (*trans*-isomer) and *iso*crotonic acid (*cis*-isomer; the *cis*-isomer is usually the less stable of the two; see §2). The use of *iso* in this connection is undesirable since it already has a specific meaning in the nomenclature of alkanes. The prefix *allo* has also been used to designate the less stable isomer (*cis*), *e.g.*, *allo*cinnamic acid.

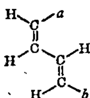
When geometrical isomers contain two or more double bonds, nomenclature may be difficult, *e g.*, VII. In this case the compound

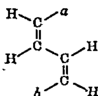


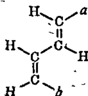
VII

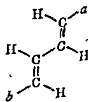
is considered as a derivative of the longest chain which contains the maximum number of double bonds, the prefixes *cis* and *trans* being placed before the numbers indicating the positions of the double bonds to describe the relative positions of the carbon atoms in the main chain; thus VII is 3-*iso*propylhexa-*cis*-2,*cis*-4-diene.

If a compound has two double bonds, *e.g*, $CHa{=}CH{-}CH{=}CHb$, four geometrical isomers are possible:







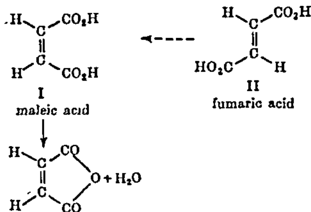


The number of geometrical isomers is 2^n, where n is the number of double bonds; this formula applies only to molecules in which the ends are different. If the ends are identical, *e.g.*, CHa=CH—CH=CHa, then the number of stereoisomers is $2^{n-1} + 2^{p-1}$, where $p = n/2$ when n is even, and $p = \frac{n+1}{2}$ when n is odd (Kuhn *et al.*, 1928).

§5. Determination of the configuration of geometrical isomers. There is no general method for determining the configuration of geometrical isomers. In practice one uses a number of different methods, the method used depending on the nature of the compound in question. The following are methods which may be used mainly for compounds that owe their geometrical isomerism to the presence of a double bond, but several of the methods are special to geometrical isomers possessing a cyclic structure (see also §7).

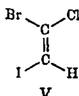
(i) **Method of cyclisation.** Wislicenus was the first to suggest the principle that *intramolecular* reactions are more likely to occur the closer together the reacting groups are in the molecule. This principle appears always to be true for reactions in which *rings* are formed, but does not hold for elimination reactions in which a double (or triple) bond is produced [see, *e.g.*, (xi)].

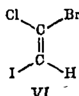
(*a*) Of the two acids maleic and fumaric, only the former readily forms a cyclic anhydride when heated; the latter does not form an anhydride of its own, but when strongly heated, gives maleic anhydride. Thus I is maleic acid, and II is fumaric acid.



Cyclisation reactions must be performed carefully, since one isomer may be converted into the other during the cyclising process, and so lead to unreliable results. In the above reaction, somewhat vigorous conditions have been used; hence there is the possibility that interconversion of the stereoisomers has occurred. Since maleic acid cyclises readily, and fumaric acid only after prolonged

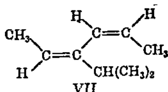
may be named *cis*-1-*bromo*-1-chloro-2-*iodo*ethylene and VI *trans*-1-*bromo*-1-chloro-2-*iodo*ethylene. This method, it must be admitted, would offer difficulties when the names are spoken.





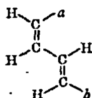
Some pairs of geometrical isomers have trivial names, *e g.*, maleic and fumaric acids, angelic and tiglic acids, etc. (*cf.* §1). Sometimes the prefix *iso* has been used to designate the *less* stable isomer, *e.g.*, crotonic acid (*trans*-isomer) and *iso*crotonic acid (*cis*-isomer; the *cis*-isomer is usually the less stable of the two; see §2). The use of *iso* in this connection is undesirable since it already has a specific meaning in the nomenclature of alkanes. The prefix *allo* has also been used to designate the less stable isomer (*cis*), *e.g.*, *allo*cinnamic acid.

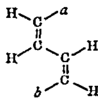
When geometrical isomers contain two or more double bonds, nomenclature may be difficult, *e.g.*, VII. In this case the compound

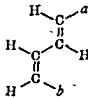


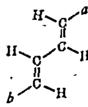
is considered as a derivative of the longest chain which contains the maximum number of double bonds, the prefixes *cis* and *trans* being placed before the numbers indicating the positions of the double bonds to describe the relative positions of the carbon atoms in the main chain; thus VII is 3-*iso*propylhexa-*cis*-2,*cis*-4-diene.

If a compound has two double bonds, *e g.*, CH*a*=CH—CH=CH*b*, four geometrical isomers are possible:







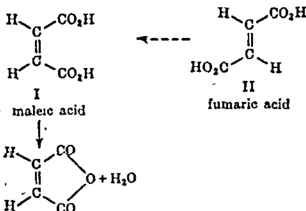


The number of geometrical isomers is 2^n, where n is the number of double bonds; this formula applies only to molecules in which the ends are different. If the ends are identical, *e.g.*, CHa=CH—CH=CHa, then the number of stereoisomers is $2^{n-1} + 2^{p-1}$, where $p = n/2$ when n is even, and $p = \frac{n+1}{2}$ when n is odd (Kuhn *et al.*, 1928).

§5. Determination of the configuration of geometrical isomers. There is no general method for determining the configuration of geometrical isomers. In practice one uses a number of different methods, the method used depending on the nature of the compound in question. The following are methods which may be used mainly for compounds that owe their geometrical isomerism to the presence of a double bond, but several of the methods are special to geometrical isomers possessing a cyclic structure (see also §7).

(i) Method of cyclisation. Wislicenus was the first to suggest the principle that *intramolecular* reactions are more likely to occur *the closer together the reacting groups are in the molecule.* This principle appears always to be true for reactions in which *rings* are formed, but does not hold for elimination reactions in which a double (or triple) bond is produced [see, *e.g.*, (xi)].

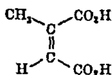
(*a*) Of the two acids maleic and fumaric, only the former readily forms a cyclic anhydride when heated; the latter does not form an anhydride of its own, but when strongly heated, gives maleic anhydride. Thus I is maleic acid, and II is fumaric acid.



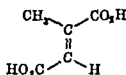
Cyclisation reactions must be performed carefully, since one isomer may be converted into the other during the cyclising process, and so lead to unreliable results. In the above reaction, somewhat vigorous conditions have been used; hence there is the possibility that interconversion of the stereoisomers has occurred. Since maleic acid cyclises readily, and fumaric acid only after prolonged

heating, the former is most probably the *cis*-isomer, and the latter the *trans* which forms maleic anhydride *via* the formation of maleic acid (see also §6). The correctness of the conclusion for the configurations of the two acids may be tested by hydrolysing maleic anhydride in the cold; only maleic acid is obtained. Under these mild conditions it is most unlikely that interconversion occurs, and so we may accept I as the configuration of maleic acid.

(*b*) Citraconic acid forms a cyclic anhydride readily, whereas the geometrical isomer, mesaconic acid, gives the same anhydride but much less readily. Thus these two acids are:

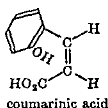


citraconic acid

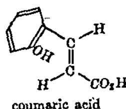


mesaconic acid

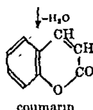
(*c*) There are two *o*-hydroxycinnamic acids, one of which spontaneously forms the lactone, coumarin, whereas the other does not. Thus the former is the *cis*-isomer, coumarinic acid, and the latter the *trans*-isomer, coumaric acid.



coumarinic acid

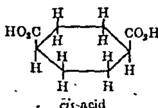


coumaric acid

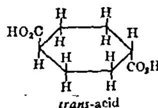


coumarin

(*d*) Two forms of hexahydroterephthalic acid are known, one of which forms a cyclic anhydride, and the other does not. Thus the former is the *cis*-isomer, and the latter the *trans* (see also §§9, 11).

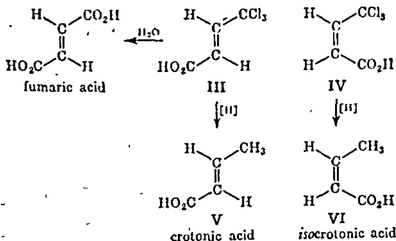


cis-acid

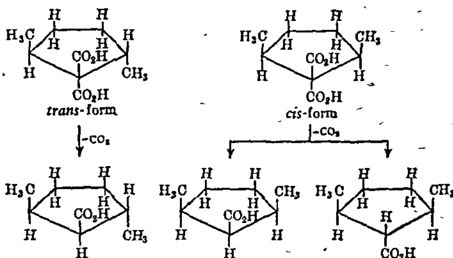


trans-acid

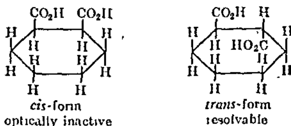
(ii) **Method of conversion into compounds of known configuration.** In a number of cases, it is possible to determine the configurations of pairs of geometrical isomers by converting them into compounds the configurations of which are already known. As an example of this type let us consider the two forms of crotonic acid, one of which is known as crotonic acid (m.p. 72°), and the other as *iso*crotonic acid (m.p. 15·5°). Now there are two trichlorocrotonic acids, III and IV, one of which can be hydrolysed to fumaric acid. Therefore this trichlorocrotonic acid must be the *trans*-isomer, III; consequently the other is the *cis*-isomer IV. Both these trichlorocrotonic acids may be reduced by sodium amalgam and water, or by zinc and acetic acid, to the crotonic acids, III giving crotonic acid, V, and IV giving *iso*crotonic acid, VI. Thus crotonic acid is the *trans*-isomer, and *iso*crotonic the *cis* (von Auwers *et al.*, 1923).



(iii) ***Method of conversion into less symmetrical compounds.*** Certain pairs of geometrical isomers may be converted into less symmetrical compounds in which the number of geometrical isomers is increased, and by considering the number of products obtained from each original stereoisomer, it is possible to deduce the configurations of the latter. *E.g.*, there are two 2.5-dimethyl*cyclo*pentane-1:1-dicarboxylic acids, and these, on heating, are decarboxylated to 2.5-dimethyl*cyclo*pentane-1-carboxylic acid. Consideration of the following chart shows that the *cis*-form of the original dicarboxylic acid can give rise to *two* stereoisomeric monocarboxylic acids, whereas *the trans-form can produce only one product.* Thus the configurations of the dicarboxylic acids are determined (see also §10).



(iv) **Method of optical activity.** In many pairs of geometrical isomers one form may possess the requirements for optical activity (§2. II), whereas the other form may not. In such cases a successful resolution of one form will determine the configuration, *e.g.*, there are two hexahydrophthalic acids; the *cis*-form possesses a plane of symmetry and consequently is optically inactive. The *trans*-form, however, possesses no elements of symmetry, and so should be resolvable; this has actually been resolved (see also §11).



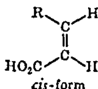
(v) **Method of dipole moments.** The unsymmetrical *cis*-form would be expected to have a large dipole moment, whereas the more symmetrical *trans*-form will have a zero or almost zero dipole moment. Results obtained by this method are in agreement with those obtained by chemical methods. Unfortunately, however, the method of dipole moments is of somewhat limited application, being confined to those cases where *a* and *b* in the molecule $Cab{=}Cab$ are simple groups, *e.g.*, alkyl and halogen. When the groups are more complex, the dipole moment of the group itself may overshadow the effects due to the *cis* and *trans* configurations of the molecule (see also §13. I).

(vi) **X-ray analysis method.** This method of determining the

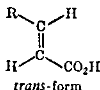
configuration of geometrical isomers is probably the best where it is readily applicable (see also §16. I).

(vii) **Ultraviolet, visible, infra-red and Raman spectra methods.** Geometrical isomers may show different spectra, *e.g.*, the intensity of the band in the ultraviolet absorption spectrum depends on the dipole moment (see Vol. I, Ch. XXXI), and this, in turn, depends on the distance between the charges. In the *trans*-form of a *conjugated* molecule, the distance between the ends is greater than that in the *cis*-form. Consequently the intensity of absorption of the *trans*-form is greater than that of the *cis* (see also §15. I). Thus, in cases such as these, it is possible to assign configurations to pairs of geometrical isomers.

(viii) **Method of surface films.** Long-chain geometrical isomers which contain a terminal group capable of dissolving in a solvent will form surface films, but only the *trans*-form can form a close-packed film, *e.g.*, the long-chain unsaturated fatty acids.

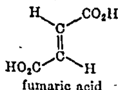


cis-form

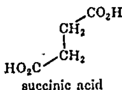


trans-form

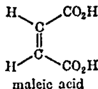
(ix) **Method of formation of solid solutions.** In compounds which owe their property of geometrical isomerism to the presence of an olefinic bond, the shape of the *trans*-form is similar to that of the corresponding saturated compound, whereas that of the *cis*-form is different, *e.g.*, the shapes of fumaric and succinic acids are similar, but the shape of maleic acid is different from that of succinic acid.



fumaric acid



succinic acid



maleic acid

Now molecules which are approximately of the same size and shape tend to form solid solutions. Thus fumaric acid forms a solid solution with succinic acid, whereas maleic acid does not; hence the configurations of maleic and fumaric acids may be determined.

(x) **Methods based on generalisations of physical properties.** Comparison of the physical properties of geometrical isomers of known configurations has led to the following generalisations:

(*a*) The melting point and intensity of absorption of the *cis*-isomer are *lower* than those of the *trans*.

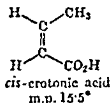
(*b*) The boiling point, solubility, heat of combustion, h
hydrogenation, density, refractive index, dipole moment
sociation constant (if the compound is an acid) of the *cis*
are *greater* than those of the *trans*.

Based on certain of these generalisations is the **Auwers**
rule (1915, 1920), *viz.*, in a pair of *cis-trans* isomers (of ali
compounds), the *cis* has the higher density and refractive i
This rule has been used to elucidate configurations, particular
terpene chemistry, *e.g.*, the menthones (see §16. VIII), but rec
it has been shown that the use of this rule may give mislea
results (see §11).

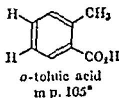
It can be seen from the above physical properties that the *tra*
form is usually the stabler of the two isomers, *i.e.*, the *trans*-isom
is the form with the lower internal energy (*cf.* §2).

Thus, in general, the above physical properties may be used t
determine the configurations of unknown geometrical isomers, but th
results should always be accepted with reserve, since exceptions are
known. Even so, determination of as many as possible of the above
physical properties will lead to reliable results, since deviations from
the generalisations appear to be manifested in only one or two
properties. It should also be noted that where the method of dipole
moments can be applied, the results are reliable [*cf.* (v)].

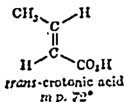
Another method based on generalisations of physical properties
is that suggested by Werner. Werner (1904) pointed out that
ethylenic *cis-trans* isomers may be compared with the *ortho-* and *para-*
isomers in the benzene series, the assumption being made that the
melting points of the *cis-* and *ortho*-isomers are lower than those of
the corresponding *trans-* and *para*-isomers, *e g.*,



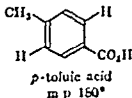
cis-crotonic acid
m.p. 15·5°



o-toluic acid
m p. 105°



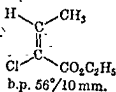
trans-crotonic acid
m p. 72°



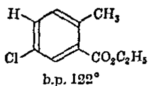
p-toluic acid
m p 180°

Thus comparison of melting points offers a means of assigning
configurations to geometrical isomers. Examination of the above

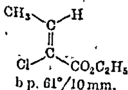
structures shows that, as far as the shape of the molecule is concerned, the benzene ring may be regarded as usurping the function of C=C in the olefinic compound. By making use of this idea, it has been possible to assign configurations to difficult cases of geometrical isomerism, *e.g.*, there are two ethyl α-chlorocrotonates, and by comparing their physical properties with ethyl 5-chloro-*o*- and 3-chloro-*p*-toluates, configurations may be assigned to the chlorocrotonates.



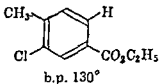
b.p. 56°/10 mm.



b.p. 122°



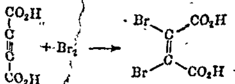
b p. 61°/10 mm.



b.p. 130°

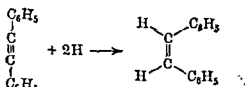
(xi) **Method of stereospecific addition and elimination reactions.** This method for determining the configurations of geometrical isomers is based on the assumption that addition reactions to a double or triple bond always occur in a definite manner—either *cis* or *trans*—for a given addendum under given conditions. Similarly, elimination reactions are also assumed to take place in a definite manner.

(*a*) **Conversion of acetylenic compounds into ethylenic compounds, and *vice versa*.** This problem was first studied by Wislicenus (1887), who suggested that when one of the acetylenic bonds is broken, the two groups of the addendum should add on in the *cis*-position, *e.g.*, the addition of bromine to acetylenedicarboxylic acid should produce dibromomaleic acid.

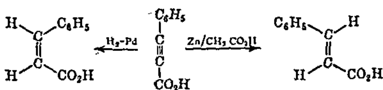


In practice, however, a mixture of dibromofumaric and dibromomaleic acids is obtained, with the former predominating. Similarly, halogen acids add on to give mainly halogenofumaric acid. Thus, in these two examples, the suggestion of Wislicenus is incorrect.

On the other hand, the reduction of tolan with zinc dust and acetic acid (Rabinovitch *et al.*, 1953) produces *iso*stilbene (the *cis*-compound):



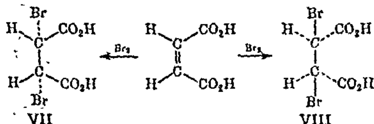
This is a *cis*-addition, but the problem of reduction of a triple bond is complicated by the fact that the results depend on the nature of the compound and the conditions used, *e.g.*, Fischer (1912) found that phenylpropiolic acid on catalytic reduction gave *cis*-cinnamic acid, whereas on reduction with zinc dust and acetic acid, *trans*-cinnamic acid was obtained.



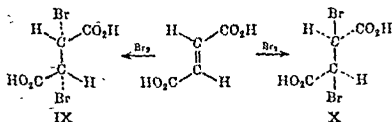
Benkeser *et al.* (1955), on the other hand, have shown that the reduction of acetylenes with lithium in aliphatic amines of low molecular weight produces *trans* olefins. It appears that, in general, chemical reduction produces the *trans* olefin, whereas catalytic hydrogenation produces the *cis* olefin. As a result of a large amount of experimental work, it has been found that addition reactions to a triple bond where the addenda are halogens or halogen acids produce predominantly the *trans*-ethylenic compound, and so, using this generalisation, one can determine the configurations of geometrical isomers when prepared from acetylenic compounds (provided, of course, the addenda are halogen or halogen acid).

Wislicenus also supposed that removal of halogen, halogen acid, etc., from olefinic compounds to produce acetylenic compounds was easier in the *cis*-position than in the *trans*. This again was shown to be incorrect experimentally, and thus the elimination reaction may be used to determine configuration if the assumption is made that *trans*-elimination occurs more readily than *cis* (see also oximes, §2f. VI).

(*b*) **Conversion of ethylenic compounds into ethane derivatives, and *vice versa*.** Just as it was *assumed* that the addition of halogens and halogen acids to a triple bond takes place in the *cis*-position, so the same assumption was made with respect to the double bond. Thus the addition of bromine to maleic acid should give *meso*-α,α'-dibromosuccinic acid. Configurations VII (formed by



attack from *behind* the molecule) and VIII (formed by attack in *front*) are identical, both being the same *meso*-dibromosuccinic acid. Similarly fumaric acid would be expected to give (±)-α:α'-dibromosuccinic acid. IX and X are mirror images, and since they will be



formed in equal amounts (see §7a. II), the racemic modification is [illegible] addition of bromine must be *trans*. In the same way it has been shown that the addition of halogen acid is also *trans*. Hence, assuming *trans* addition always occurs with these addenda, the nature of the products indicates the configuration of the ethylenic compound.

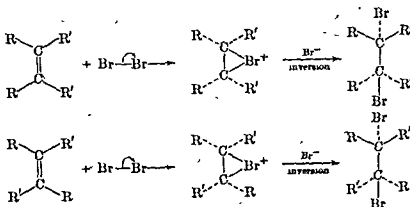
The configuration of the product formed by hydroxylation of a double bond depends on the nature of the hydroxylating agent used and on the conditions under which the reaction is carried out. Permanganate and osmium tetroxide apparently always give *cis* addition, whereas permonosulphuric acid (Caro's acid) and perbenzoic acid give *trans* addition. On the other hand, hydroxylation

Reagent	Type of addition	Maleic acid	Fumaric acid
$KMnO_4$	*cis*	*meso*tartaric acid	DL-tartaric acid
OsO_4	*cis*	*meso*tartaric acid	DL-tartaric acid
H_2SO_5	*trans*	DL-tartaric acid	*meso*tartaric acid
C_6H_5·CO·O_2H	*trans*	DL-tartaric acid	*meso*tartaric acid
H_2O_2—OsO_4	*cis*	*meso*tartaric acid	DL-tartaric acid
H_2O_2—SeO_2	*trans*	DL-tartaric acid	*meso*tartaric acid

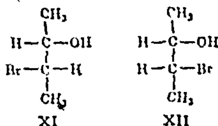
with hydrogen peroxide catalysed by osmium tetroxide in *tertiary*-butanol gives *cis* addition ; if the reaction is catalysed by selenium dioxide in *tertiary*-butanol or in acetone, then the addition is *trans* (see also below). The table above shows the products formed by hydroxylation of maleic and fumaric acids.

Mechanism of stereochemical additions. The mechanisms of the stereochemical additions to a double and a triple bond are still not certain. The addition of halogen is *trans.* If the two halogen atoms approach on the same side of the attacked molecule and add *simultaneously* to the two carbon atoms, *i.e.*, we have a four-centre side-approach reaction (§3b. III), the resulting addition must be *cis*. If the two halogen atoms add on *singly*, then *cis* or *trans* addition could result Since *trans* addition occurs in practice, the addition must occur by a two-step mechanism, *i.e.*, the halogen atoms must add on one at a time (*cf.* ethylene, Vol. I). It should be noted that *trans* addition corresponds to the "inversion" of *one* of the two carbon atoms linked by the double bond.

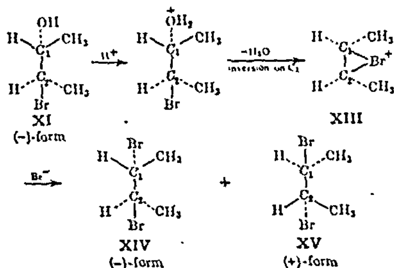
Roberts and Kimball (1937) suggested that the addition to a double bond occurs *via* the formation of a *cyclic* halogenium ion, *e.g.*,

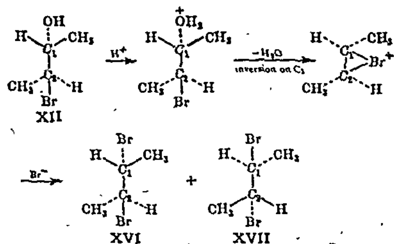


well at either carbon atom). Winstein and Lucas (1939) have demonstrated the existence of the above planar intermediate in the following way. These authors studied the action of fuming hydrobromic acid on optically active *threo*-3-bromo-butan-2-ol, XI, and found that the (−)-form gave (±)-2:3-dibromobutane. This result can only be explained by assuming the intermediate formation of the 1:2-dimethylbromonium ion, XIII. If the bromine ion attacks



C_1 of XIII, then the configuration of C_1 is inverted *again*, and C_2 retains its original configuration; this results in the formation of XIV, the (−)-form corresponding to the configuration of XI, *i.e.*, XIV is formed by *complete* retention. On the other hand, if the

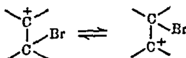




bromine ion attacks C_2, then the configurations of both C_1 and C_2 are inverted, *i.e.*, there is *complete* inversion; this results in the formation of XV, the enantiomorph of XIV. Since both forms will be produced in equal amounts (§7a. II), the product is the racemic modification, the result actually obtained by Winstein and Lucas.

The above mechanism also explains the formation of *meso*-2:3-dibromobutane by the action of fuming hydrobromic acid on optically active *erythro*-3-bromobutan-2-ol, XII. XVI and XVII are identical and correspond to the *meso*-form.

It might be noted, in passing, that it has been suggested that the intermediate bromonium ion is a resonance hybrid.



An alternative explanation for the addition of bromine to an olefinic compound is as follows. At first the Br^+ ion is held loosely by the π-electrons to form a **π-complex** (Fig. 4 *a*). The molecule is planar at this moment. Then tetrahedral hybridisation of C_1

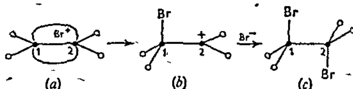
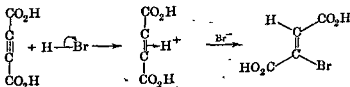


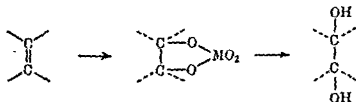
FIG. 4 4.

occurs, C_2 still retaining its trigonal hybridisation (Fig. 4 *b*). The bromine ion now attacks C_2, but does so on the side *remote* from the bromine atom on C_1, partly because of the repulsion of the bromine atom already on C_1, and partly because of the hindering effect of this bromine atom due to its actual bulk (spatial factor; see Vol. I). When C_2 is attacked on the "under" side, it hybridises to the tetrahedral form, and does so with "inversion" (*i.e.*, in the opposite sense to C_1; Fig. 4 *c*).

Addition of halogen to a triple bond may be assumed to take place by mechanisms similar to those suggested for a double bond, *e g.*, the addition of hydrogen bromide to acetylene dicarboxylic acid to form bromofumaric acid may be formulated:

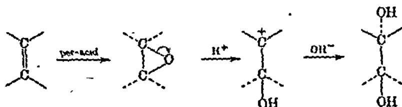


Now let us consider the mechanism of hydroxylation, *i.e.*, the addition of two hydroxyl groups to a double bond. With potassium permanganate and osmium tetroxide the *cis*-addition is readily explained by assuming the formation of a cyclic organo-metallic intermediate.



This cyclic intermediate is definitely known in the case of osmium tetroxide (see Vol. I); for potassium permanganate it may be assumed that the permanganate ion, MnO_4^- (or the manganate ion, $MnO_4^{=}$), behaves in a similar manner.

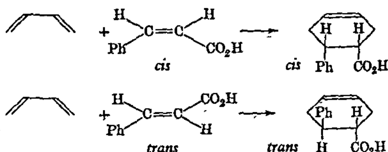
With per-acids the hydroxylation results in *trans*-addition. The first product of oxidation is an epoxide (Prileschaiev reaction; see Vol. I). This compound may then be assumed to be hydrolysed with a Walden inversion occurring at *one* of the two carbon atoms due to the attack of a hydroxyl ion from a position remote from the one already present (*cf.* the addition of bromine).



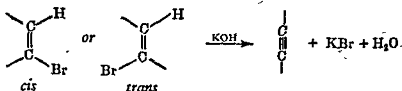
The addition of hydrogen peroxide may result in *cis* or *trans* compounds. Which occurs depends on the conditions of the experiment, *e.g.*, the catalyst (see above). Where *trans*-addition occurs, the mechanism may possibly be through the epoxide, but a free hydroxyl radical mechanism could also result in the *trans*-glycol. *Cis*-addition in the presence of certain oxides probably occurs *via* a cyc c intermediate.

The addition of a dienophile to a diene in the Diels-Alder reaction is stereospecific; *cis* addition always occurs (see Vol. I). Since it is usually possible to determine the configuration of the cyclic adduct, this offers a means of ascertaining the configuration of the dienophile. *E.g.*, butadiene forms adducts with *cis*- and *trans*-cinnamic acids, and hence determination of the configurations of

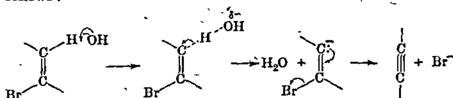
the stereoisomeric adducts will determine the configurations of the cinnamic acids (see §11); thus:



Mechanism of stereochemical eliminations. In the elimination reaction:

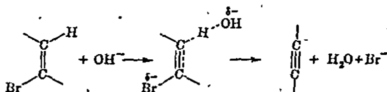


the elimination has been found to occur more readily in the *trans*-isomer than in the *cis*. This may be explained by assuming that the reaction takes place by removal of each atom *one at a time* as follows:

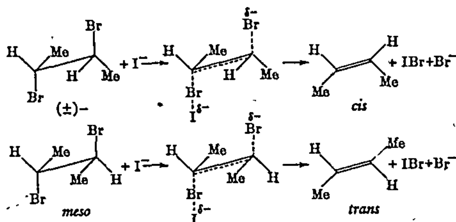


The hydroxyl ion can more easily approach the hydrogen atom when the latter is in the *trans*-position with respect to the bromine atom because the repulsion between the negative hydroxyl ion and the negatively charged bromine atom is less for the *trans* molecule than for the *cis* (*cf.* the Walden inversion, §4. III).

On the other hand, many believe that the elimination occurs by a concerted (*i.e.*, simultaneous) process in which the transition state is planar:



Now let us consider eliminations in ethane derivatives to form ethylene derivatives, *e.g.*, the debromination of 2:3-dibromobutane by means of potassium iodide. According to Hughes and Ingold, such eliminations (bimolecular elimination reactions, E2) take place more readily when the four centres involved lie in one plane. Thus, in the transition state, the two carbons (of the CBr groups) and the two bromine atoms will all lie in the same plane and at the same time the two bromine atoms will be in the staggered position. Now 2:3-dibromobutane exists in (+)-, (−)-, and *meso*-forms, and it has been shown that the (±)-form gives *cis*-butene, whereas the *meso*-form gives *trans*-butene. These eliminations may therefore be written as follows (following Winstein *et al.*, 1939; the iodine atom is probably in the same plane as the other four groups involved in the planar transition state):

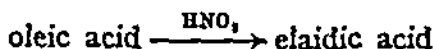


In the (±)-form, as the transition state changes into the ethylene compound, the two methyl groups become eclipsed; in the *meso*-form a methyl group becomes eclipsed with a hydrogen. Thus the energy of activation of the transition state of the (±)-form will be greater than that of the *meso*-form and consequently the latter should be formed more readily, *i.e.*, the *meso*-form should undergo debromination more readily than the (±)-form. Young *et al.* (1939) have shown that this is so in practice, the rate of debromination being about twice as fast. These authors also showed that the rate of debromination of *meso*-stilbene dibromide (Ph·CHBr·CHBr·Ph) is about one hundred times as fast as that of the (±)-form.

§6. Interconversion (stereomutation) of geometrical isomers. The *cis*-isomer, being usually the more labile form, is readily converted into the *trans*-form under suitable physical or

chemical conditions. The usual chemical reagents used for stereomutation are halogens and nitrous acid, *e.g.*,

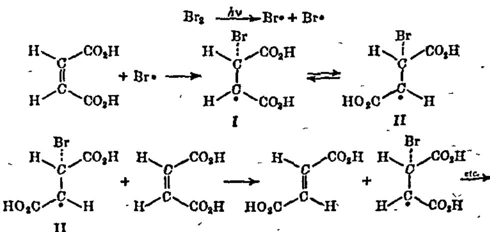




Other methods such as distillation or prolonged heating above the melting point also usually convert the *cis*-isomer into the *trans*, but, in general, the result is a mixture of the two forms.

The conversion of the *trans*-isomer into the *cis* may be effected by means of sunlight, but the best method is to use ultraviolet light in the presence of a trace of bromine.

Many theories have been proposed for the interconversion of geometrical isomers, but none is certain. To effect conversion, the double bond must be "dissociated" so as to allow rotation about the single bond (*i.e.*, the σ-bond; see §3). Let us consider the conversion of maleic acid into fumaric acid under the influence of light and in the presence of a trace of bromine. One mechanism that has been suggested for this change is a free-radical chain reaction, since the conversion does not appear to be effected by bromine in the dark. Thus:

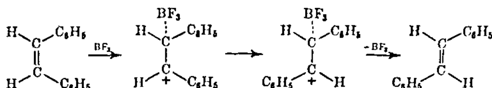


In free radicals I and II, the upper carbon atom is in a state of tetrahedral hybridisation, and the lower one (the free radical part) in a trigonal state (and therefore flat). Owing to the repulsion between the carboxyl groups, configuration I tends to change into configuration II by rotation about the single bond (*cf.* §4. II). If II now reacts with a molecule of maleic acid, the latter is converted into a free radical containing the bromine atom, and II is converted

into fumaric acid *if "inversion" occurs on the lower carbon atom*, if no "inversion" occurs, II would form maleic acid again.

Similarly, various other reagents are also believed to act by a free-radical mechanism, *e.g.*, the conversion of *cis*-stilbene into *trans*-stilbene by means of light in the presence of hydrogen bromide. In the absence of light, the conversion takes place very slowly, but in the presence of oxygen or benzoyl peroxide, the conversion is rapid. These reagents are known to generate free radicals, this supports the free-radical mechanism, the reaction being initiated by the formation of free radicals from the hydrogen bromide. Furthermore, if the reaction is carried out in the presence of benzoyl peroxide *and quinol*, the conversion of *cis*- into *trans*-stilbene is extremely slow. This is in keeping with the free-radical mechanism, since it is known that quinol removes free radicals.

Boron trifluoride also catalyses the conversion of *cis*- into *trans*-stilbene. In this case the mechanism is less certain, but a reasonable one is:

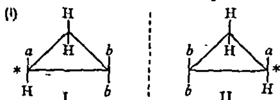


Now let us consider *thermal* interconversion. Kistiakowsky (1935) has shown experimentally that there are at least two mechanisms for thermal *cis-trans* isomerisation of ethylene compounds, and that both are first-order reactions. Experimental results have also shown that one mechanism requires a high and the other a low energy of activation. In the transition state (in both thermal and chemical isomerisations), the two parts of the molecule are perpendicular to each other. To reach this state the double bond, as we have seen, must undergo "dissociation", this occurs by the decoupling of the π-electrons. The spins of these electrons may remain anti-parallel in the perpendicular (*i.e.*, transition) state. This type of "dissociation" of a double bond requires energy of about 40 kg cal., and the transition is said to be from a singlet ground state to an upper singlet state. On the other hand, it is also possible for the spins of the π-electrons to be parallel (this state is said to be the triplet state), and the energy required for this "dissociation" is about 25 kg cal.

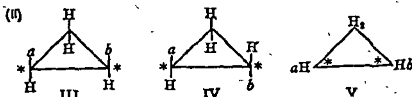
§7. STEREOCHEMISTRY OF CYCLIC COMPOUNDS

Geometrical and optical isomerism may exist in any sized ring. In the following account, the saturated rings are treated as rigid flat structures, and the groups attached to the ring-carbon atoms are regarded as being above or below the plane of the ring (see also, in particular, *cyclo*hexane compounds, §11). Furthermore, the examples described deal *only* with those cases in which the asymmetric carbon atoms are part of the saturated ring system. In general, the pattern of optical isomerism followed by cyclic compounds is similar to that of the acyclic compounds. The main difference between the two is that, since there is no free rotation about ring-carbon atoms, geometrical isomerism may therefore be manifested as well as optical isomerism. On the other hand, geometrical isomerism may exist without optical isomerism (see §5 for methods of determination of the configuration of geometrical isomers; see also §§9, 10, 11).

§8. *cyclo*Propane types. Molecule I contains one asymmetric carbon atom (*), and is not superimposable on its mirror image molecule II. Thus I and II are enantiomorphs, *i.e.*, a *cyclo*propane

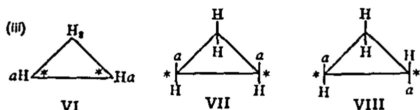


derivative containing one asymmetric carbon atom can exist in two optically active forms (and one racemic modification; *cf.* §7a. II). Molecule III contains two different asymmetric carbon atoms, and since it has no elements of symmetry (§6. II), it is not superimposable on its mirror image molecule. Thus *III* can exist in two *optically*

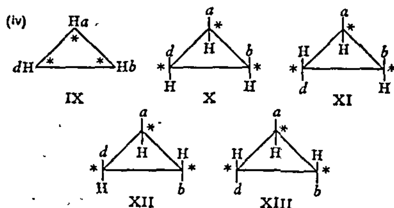


active forms (and one racemic modification). Structure III, however, is capable of exhibiting geometrical isomerism, the two geometrical isomers being III and IV. Now IV also contains two different asymmetric carbon atoms, and these are not disposed towards each other as in III. Since IV possesses no elements of symmetry, it can also exist in two optically active forms which are

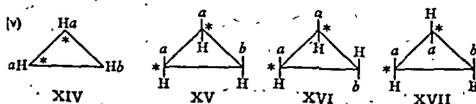
different from those of III. Thus V, which may be regarded as the non-committal way of writing the configurations III and IV, is similar, as far as *optical isomerism* is concerned, to the acyclic molecule *Cabd·Cabe*, *i.e.*, there are four optically active forms in all (two pairs of enantiomorphs). In general, any monocyclic system can exist in 2^n optically active forms, where n is the number of different asymmetric ring-carbon atoms (*cf.* §7c. II). Molecule VI



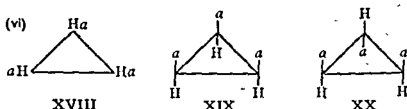
contains two similar asymmetric carbon atoms, and can exist as geometrical isomers VII and VIII. VII has a (vertical) plane of symmetry and therefore represents a *meso*-form. VIII, however, possesses no elements of symmetry and can therefore exist in two optically active forms (and one racemic modification). IX contains



three different asymmetric carbon atoms and can therefore exist in $2^3 = 8$ optically active forms (four pairs of enantiomorphs). Each pair of enantiomorphs is derived from the *four* geometrical isomers X–XIII. Inspection of these configurations shows that all of them possess no elements of symmetry. XIV contains two similar



asymmetric carbon atoms, and the third carbon atom is pseudoasymmetric (*cf.* §7d. II). Three geometrical isomers, XV–XVII, are possible; XV and XVI each possess a (vertical) plane of symmetry, and therefore each represents a *meso*-form. XVII, however, possesses no elements of symmetry and so can exist in two optically active forms (and one racemic modification). XVIII contains three similar asymmetric carbon atoms which are all pseudo-asymmetric. Two geometrical isomers are possible, XIX and XX, both of which

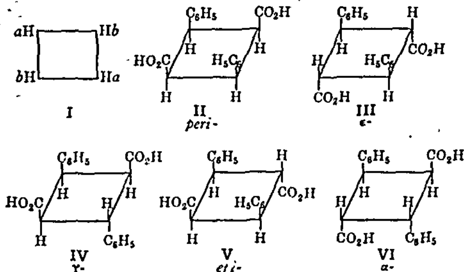


possess at least one (vertical) plane of symmetry, and therefore represent *meso*-forms.

In the above account, the stereochemistry of the *cyclo*propane ring has been dealt with from the theoretical point of view, and thus most of the ideas connected with the stereochemistry of monocyclic systems have been described. In the following sections more emphasis is laid on specific examples, and any further points that arise are dealt with in the appropriate section.

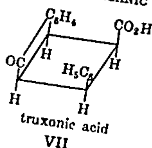
§9. *cyclo*Butane types. Two important examples of the *cyclo*butane type are truxillic and truxinic acids; truxillic acid is 2:4-diphenyl*cyclo*butane-1:3-dicarboxylic acid, and truxinic acid is 3:4-diphenyl*cyclo*butane-1:2-dicarboxylic acid. *cis*-Cinnamic acid (allocinnamic acid), *on irradiation with light, forms mainly* β-truxinic acid and *trans*-cinnamic acid, together with some of the dimer of the latter, α-truxillic acid (de Jong, 1929). Bernstein *et al.* (1943) found that irradiation of commercial *trans*-cinnamic acid gave only β-truxinic acid. When *trans*-cinnamic acid was slowly recrystallised from aqueous ethanol, dried, and then irradiated, only α-truxillic acid was obtained. Truxillic and truxinic acids have been isolated from natural sources.

Truxillic acid. This acid can exist theoretically in five stereoisomeric forms, all of which are known (the acid is of the type I). All five are *meso*-forms, II–V having planes of symmetry, and VI a centre of symmetry. The configurations of these stereoisomers have been assigned as follows. When *one* of the carboxyl groups is converted into the anilido-group, $\cdot CONH\cdot C_6H_5$, *two* of the five forms give optically active compounds, each giving a pair of

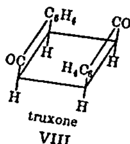


enantiomorphs. Now only the stereoisomers with the two phenyl groups in the *trans*-position can produce asymmetric molecules under these conditions; the remaining forms will each have a (vertical) plane of symmetry. Thus only IV and VI satisfy the necessary conditions. One of these is known as the α-acid (m.p. 274°) and the other the γ-acid (m.p. 288°). This then raises the problem: Which is which? This is readily answered by the fact that of the anilido-derivatives of these two acids, only one can be dehydrated to a cyclic *N*-phenyl imide, —CO—N(C_6H_5)—CO—. This reaction can be expected to take place only when the two carboxyl groups are in the *cis*-position (see §5. i). Therefore IV is γ-truxillic acid, and VI *is α-truxillic acid (since the acid with the melting point 288° has been called the* γ-acid). By considering the ease of formation of the cyclic anhydride, the configurations of the remaining three stereoisomers may be determined. Two form anhydrides readily, and therefore one of these acids must be II and the other III. The third acid does not form its own anhydride, but gives a mixture of the anhydrides produced by II and III. Thus the third acid, *epi*-truxillic acid, is V. The final problem is to decide which of the two, II and III, is *peri*-truxillic acid, and which is ε-truxillic acid. *peri*-Truxillic acid, under the influence of aluminium chloride, undergoes an internal Friedel–Crafts reaction to form a truxonic acid, VII, and a truxone, VIII. This is only possible when the phenyl and carboxyl groups are in the *cis*-position: (overleaf). Thus II is *peri*-truxillic acid, and therefore III is ε-truxillic acid.

Truxinic acid. This acid can exist theoretically in six geometrical isomeric forms, four of which are resolvable; thus ten forms in all are possible theoretically. Truxinic acid is of

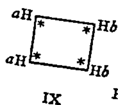


truxonic acid
VII

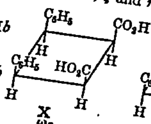


truxone
VIII

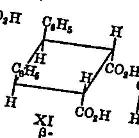
the type IX, and the six geometrical isomers possible are X–XV. X and XI are *meso*-forms (each has a plane of symmetry); XII–XV are resolvable (theoretically), since all possess no elements of symmetry. The configurations of these stereoisomers have been determined by methods similar to those used for the truxillic acids; it appears, however, that only four of these six forms are known with certainty, *viz.*, β, δ, ζ and *neo*.



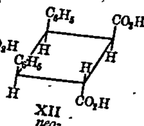
IX



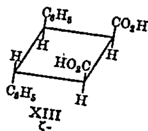
X
ω-



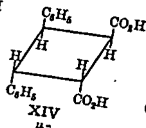
XI
β-



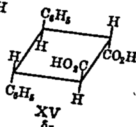
XII
neo-



XIII
ζ-

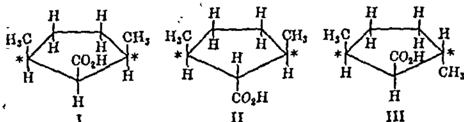


XIV
μ-

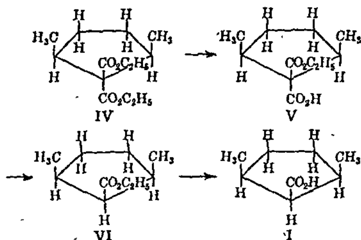


XV
δ-

§10. *cyclo*Pentane types. A number of examples involving the stereochemistry of the five-membered ring occur in natural products, *e.g.*, camphoric acid (§23a. VIII), furanose sugars (§7b. VII). In this section we shall discuss the case of 2:5-dimethyl*cyclo*pentane-1:1-dicarboxylic acid. This acid can exist in two geometrical isomeric forms, which may be differentiated by decarboxylation, the *cis*-isomer giving two monocarboxylic acids, I and II, and the *trans*-isomer one monocarboxylic acid, III (see §5. iii). All three acids contain two similar asymmetric carbon atoms and one pseudo-asymmetric carbon atom. Both I and II possess a (vertical) plane of symmetry, and are therefore *meso*-



forms; III possesses no elements of symmetry, and can therefore exist in two optically active forms (and one racemic modification). All the possible forms are known, and I and II have been differentiated as follows. The diethyl ester of the *cis*-dicarboxylic acid, IV, can be partially hydrolysed to the monoethyl ester, which most probably has the configuration V. This is based on the assumption that the carbethoxyl group on the same side as the two methyl groups is far more resistant to attack than the other carbethoxyl group because of the spatial effect (see Vol. I). Decarboxylation of V gives VI, and this, on hydrolysis, gives I. Thus the configuration of I (and therefore also of II) is determined.



The above treatment of the *cyclo*pentane derivatives has been based on the assumption that the ring is planar. This classical treatment leads to agreement between prediction and the number of stereoisomers actually obtained (see *cyclo*hexane, §11, for a further discussion of this problem). It is now known that the *cyclo*pentane ring is not planar; the puckering, however, is very small. The non-planarity of this ring has been shown from entropy determinations (Aston *et al.*, 1941), spectroscopic studies (Miller *et al.*, 1950), and from a study of the polarisabilities of C—$C_{aliphatic}$ and C—H bonds (Le Fèvre *et al.*, 1956).

§11. cyclo-Hexane types. The stereochemistry of *cyclo*hexane and its derivatives presents a detailed example of the principles of conformational analysis (§4a. II). On the basis of the tetrahedral theory, two forms are possible for *cyclo*hexane, neither of which is planar. These two forms, known as **boat and chair conformations** (Fig. 5), were first proposed by Sachse (1890 ; see Vol. I, Ch. XIX), who also pointed out that both are strainless. Hassel *et al.* (1943)

"boat" or C form

"chair" or Z form

FIG. 4.5.

showed by means of electron diffraction studies that at room temperature most of the molecules existed mainly in the chair conformation. Pitzer (1945) then showed by calculation that the energy difference between the two forms is about 5·6 kg.cal./mole (the boat form having the higher energy content ; see also below). This value, however, is too small for stability, and consequently neither conformation retains its identity, each being readily converted into the other.

Although these two forms are free from "angle strain", forces due to steric repulsion (*i.e.*, repulsive forces between non-bonded atoms) are acting, and it is because of their different total effects that the two conformations differ in energy content. A simple method of calculating this energy difference has been introduced by Turner (1952). Fig. 6 (*a*) and 6 (*b*) represent the chair and boat conformations and the directions of the C—H bonds. In the

(*a*) chair form

(*b*) boat form

FIG. 4.6.

chair conformation, all the C—H bonds on adjacent carbons are in the skew position (*i.e.*, the arrangement is skew as in the skew form of *n*-butane, §4. II). On the other hand, in the boat conformation there are four skew interactions (1:2, 3:4, 4:5, and 6:1)

and two eclipsed interactions (2:3 and 5:6). According to Pitzer (1940), skew interaction of the hydrogens in *n*-butane is 0·8 kg.cal., and an eclipsed interaction is 3·6 kg.cal. Thus the steric strain in the chair form is $6 \times 0{\cdot}8 = 4{\cdot}8$ kg.cal., and in the boat form $4 \times 0{\cdot}8 + 2 \times 3{\cdot}6 = 10{\cdot}4$ kg.cal. Thus the boat form has the greater energy content, and the amount (according to the above method of calculation) is 5·6 kg.cal. There is, however, a further interaction in the boat form, *viz.* the interaction of the two flagpole hydrogens (at positions 1 and 4). These are closer together than any other two hydrogens (see table below) and so produce an additional steric repulsion. The actual value of this interaction is not certain, but it is believed to be about the same as that of two eclipsed hydrogens. Thus the energy content of the boat form is $10{\cdot}4 + 3{\cdot}6 = 14$ kg.cal., and hence the boat form contains $14 - 4{\cdot}8 = 9{\cdot}2$ kg.cal. more than the chair form.

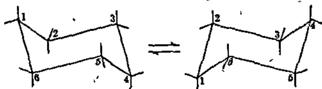
Inspection of Fig. 6 (*a*) shows that the twelve hydrogen atoms in the chair conformation are not equivalent; there are two sets of six. In one of these sets the six C—H bonds are parallel to the threefold axis of symmetry of the molecule; these are the **axial** (*a*) **bonds** (they have also been named ε- or *polar* bonds). In the other set the six C—H bonds make an angle of 109° 28′ with the axis of the ring (or ± 19° 28′ with the horizontal plane of the ring); these are the **equatorial** (*e*) **bonds** (they have also been named κ-bonds). On the other hand, in Fig. 6 (*b*) it can be seen that the "end" of the boat is different stereochemically from the chair conformation; the various C—H bonds have been named:—**flagpole** (*fp*), **bowsprit** (*bs*), **boat-equatorial** (*be*), and **boat-axial** (*ba*). Angyal and Mills (1952) have calculated the distances between the various hydrogen atoms (and carbon atoms) in both the chair and boat conformations.

Conformation	Position	H—H (A)
Chair (Fig. 6 *a*)	1*e*:2*e*	2·49
	1*e*:2*a*	2·49
	1*a*:2*a*	3·06
	1*a*:3*a*	2·51
Boat (Fig. 6 *b*)	2*a*:3*a*	2·27
	2*e*:3*e*	2·27
	1*fp*:4*fp*	1·83

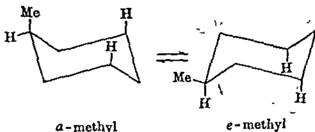
It appears that the boat conformation occurs in relatively few cases, and so in the following account we shall only study the problem of the chair conformation. Inspection of the above table shows

that a 1:2-interaction *for two* adjacent *equatorial hydrogens or for* an equatorial and adjacent axial hydrogen is about the same as for a 1:3-interaction for two *meta* axial hydrogens. Furthermore, a study of accurate scale models has shown that with any axial substituent (which is necessarily larger than hydrogen), the 1:3-interactions are larger than the 1:2-interactions when the same substituent is equatorial. Using these principles, we can now proceed to study the conformations of *cyclo*hexane derivatives.

Because of the flexibility of the chair conformation, one chair form can readily pass through a plane to form the other chair form, and in doing so all *a*- and *e*-bonds in the first now become *e*- and *a*-bonds, respectively, in the second.



Both forms are identical and so cannot be distinguished. If, however, one hydrogen is replaced by some other atom or group, the two forms are no longer identical, *e.g.*, methyl*cyclo*hexane. *In the*



a - methyl *e* - methyl

a-methyl conformation there are 1:3-interactions acting, whereas in the *e*-methyl conformation these interactions are absent; instead, the *weaker* 1:2-interactions are acting. Thus the energy content of axial conformation is greater than that of the equatorial, and consequently the latter will be the preferred form. Hassel (1947) has shown experimentally from electron-diffraction studies that the *e*-methyl conformation predominates in methyl*cyclo*hexane. Hassel *et al.* (1950) have also shown that in chloro*cyclo*hexane the *e*-form also predominates and that very little of the *a*-form is present.

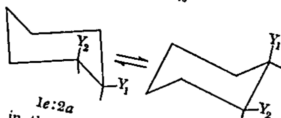
Now let us discuss the conformations of disubstituted *cyclo*hexanes. Here we have a number of factors to consider: *position isomerism, stereoisomerism (geometrical and optical), the relative* sizes of the two substituents, and the nature of the substituents.

(i) **1:2-Compounds**

Classical formula *Conformations*



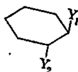
cis-1:2



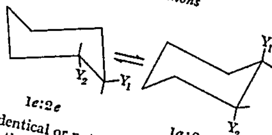
It should be noted that in these *cis*-compounds one substituent must be axial and the other equatorial. If the substituents differ in size, the 1:3-interactions will be most powerful when the larger group is axial. Thus the conformation with the lower energy will be the one in which the larger group is equatorial, *i.e.*, this is the preferred form. An example of this type is *cis*-2-methyl*cyclo*hexanol; the methyl group is larger than the hydroxyl, and so the preferred form can be expected to be 1*a*-hydroxyl-2*e*-methyl. This has been shown to be so in practice. In general, the greater the difference in size between the two substituents, the greater will be the predominance of the form with the larger group in the equatorial conformation.

The classical formula of the *cis*-compound when the two substituents are identical has a plane of symmetry and is therefore not resolvable. On the other hand, the two conformations are mirror images but not superimposable and hence, in theory, are resolvable. Such compounds, however, have never yet been resolved. The reason for this is that the two forms are separated by such a low energy barrier that they are readily interconvertible.

Classical formula *Conformations*



trans-1:2

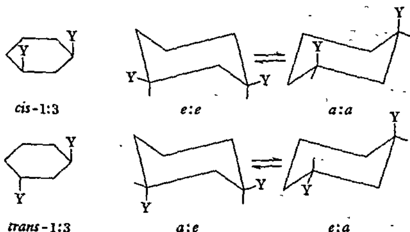


Whether Y_1 and Y_2 are identical or not, the two conformations are different, and because of the 1:3-interactions the *e:e*-form will be the preferred form. Furthermore, this form will be more stable than the *cis*-isomer (*a·e*-form). An example that illustrates this is 2-methyl*cyclo*hexanol. The *trans*-form has been shown to be more

stable than the *cis*-; the latter is readily converted into the former when heated with sodium, and also the reduction of 2-methyl*cyclo*hexanone (with sodium and ethanol) produces the *trans*-alcohol.

Both the classical formula and the *e:e*- (and *a:a*) conformation of the *trans*-1:2-compound (whether Y_1 and Y_2 are identical or not) are not superimposable on their mirror images and hence should be optically active. This has been found to be so in practice.

(ii) **1:3-Compounds**

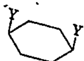


The two *trans*-conformations are identical when the two Y groups are identical. The *cis-e:e*-form will be more stable than the *cis-a:a*, and will also be more stable than the *trans-e:a*-conformation, *e.g.*, the most stable conformation of 1:3-dimethyl*cyclo*hexane has been shown to be the *cis*-1:3-*e:e*-form. It should be noted that this situation is the reverse of that of the 1:2-dimethyl*cyclo*hexanes.

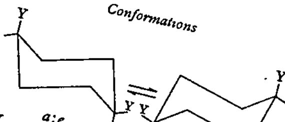
The Auwers-Skita rule (§5 (x)*b*) has been shown to break down when applied to 1:3-disubstituted *cyclo*hexanes: the reverse holds good. Allinger (1954) modified the rule for *cyclo*hexanes as follows: The isomer which has the higher boiling point, refractive index and density is the one with the less stable configuration. Thus, according to this rule, the *trans*-1:3-disubstituted *cyclo*hexanes have the higher physical constants (the *trans*-form has more axial substituents than the more stable *cis*-form); *e.g.*, Macbeth *et al.* (1954) have shown that the physical constants of (±)-*trans*-3-methyl*cyclo*hexylamine are higher than those of its *cis*-isomer.

(iii) **1:4-Compounds**

Classical formula *Conformations*



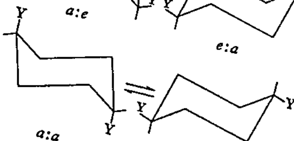
cis-1:4



a:e e:a



trans-1:4



a:a e:e

The two *cis*-conformations are identical when the Y groups are identical. Also, the *trans-e:e*-form will be more stable than the *cis-a:e*-form.

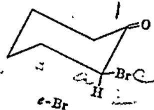
The arguments used for the disubstituted *cyclo*hexanes can also be applied to the higher substituted *cyclo*hexanes. As the result of a large amount of work, the following generalisations may be made:

(i) In *cyclo*hexane systems, mono-, di-, tri-, and poly-substituted derivatives always tend to take up the chair conformation whenever possible.

(ii) The chair conformation with the maximum number of equatorial substituents will be the preferred conformation. This generalisation, however, is only satisfactory when the internal forces due to dipole interactions or hydrogen bonding are absent. When these are present, it is necessary to determine which forces predominate before a conformation can be assigned to the molecule. As an illustration of this problem, we shall consider 2-bromo*cyclo*hexanone; the two possible chair forms are:



a-Br



e-Br

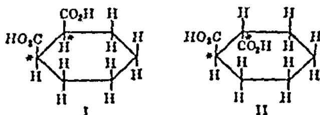
On the basis that a substituent preferably takes up an equatorial conformation, it would therefore be expected that the conformation

2*e*-bromo*cyclo*hexanone would be favoured. Infrared studies, however, have shown that the *a*-bromo conformation predominates. This has been explained as follows. The C—Br and C=O bonds are both strongly polar, and when the bromine is equatorial the dipolar repulsion is a maximum, and a minimum when the bromine is axial. Since the axial form predominates, this equatorial dipolar repulsion must therefore be larger than the 1:3-interactions. When, however, other substituents are present, the 1:3-interactions may become so large as to outweigh the dipolar effect and the bromine would now be equatorial. Such is the case with 2-bromo-4:4-dimethyl*cyclo*hexanone.



(iii) The energy barriers between the various conformations (of the cases studied so far) are too small to prevent interconversion. Up to the present time, the number of geometrical (and optical) isomers obtained from a given *cyclo*hexane derivative is in agreement with the number that can be expected from a planar ring with the substituents lying above and below the plane of the ring. We shall now, therefore, discuss the stereochemistry of some *cyclo*hexane derivatives from the classical point of view.

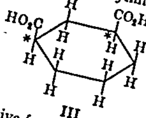
(i) *Hexahydrophthalic acids* (*cyclo*hexane-1:2-dicarboxylic acids). Two geometrical isomers are theoretically possible, the *cis*, I, and



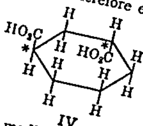
the *trans*, II. Molecule I has a plane of symmetry, and therefore represents the *meso*-form; II has no elements of symmetry, and can therefore exist in two optically active forms (and one racemic modification). All of these possible forms are known, and it has been found that the *cis*-compound, I, forms a cyclic anhydride readily, whereas the *trans*-compound, II, forms a cyclic anhydride with difficulty (*cf.* §5. i).

(ii) *Hexahydroisophthalic acids* (*cyclo*hexane-1:3-dicarboxylic acids). Two geometrical isomers are possible; the *cis*-form, III,

has a plane of symmetry, and therefore represents the *meso*-form; IV has no elements of symmetry, and can therefore exist in two



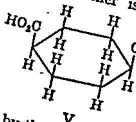
III



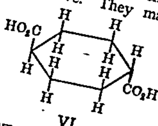
IV

optically active forms (and one racemic modification). All of these forms are known; the *cis*-isomer forms a cyclic anhydride, whereas the *trans*-isomer does not.

(iii) *Hexahydroterephthalic acids* (*cyclo*hexane-1:4-dicarboxylic acids). Two geometrical isomers are possible; the *cis*-form, V has a plane of symmetry, and the *trans*-form, VI, a centre of symmetry. Hence neither is optically active. They may be dis-



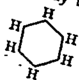
V

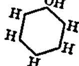


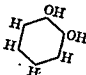
VI

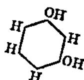
tinguished by the fact that the *cis*-isomer forms a cyclic anhydride, whereas the *trans*-isomer does not.

(iv) *Inositol* (hexahydroxy*cyclo*hexane). There are eight geometrical isomers possible theoretically, and only *one* of these is not superimposable on its mirror image molecule; thus there are nine forms in all (and also one racemic modification). If we imagine that we are looking down at the molecule, and insert the groups which appear *above* the plane of the ring, then the eight geometrical isomers may be represented as follows:

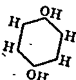


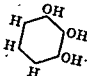


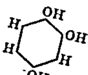




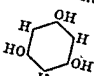
meso-inositol







resolvable



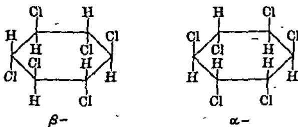
scyllitol

Examination of these configurations shows that all except one—the one labelled resolvable—have at least one plane of symmetry, and so are all *meso*-forms. All the *meso*-forms and both of the optically active forms are known; of these *meso*-inositol, scyllitol and (+)- and (−)-inositol occur naturally. The optically active inositols are examples of centro-asymmetric compounds (§6. II).

(v) *Benzene hexachloride* (hexachloro*cyclo*hexane). Here again eight geometrical isomers are possible theoretically; seven are known, α, β, γ, δ, ε, η, θ; the γ-isomer is a powerful insecticide (see Vol. I). All have been shown to exist in the chair form, and the conformations that have been assigned are:

α-, aaeeee; β-, eeeeee; γ-, aaaeee; δ-, aeeeee; ε-, aeeaee.

Of these forms, it is the β- which loses hydrogen chloride with the greatest difficulty (Cristol *et al.*, 1951). This is in keeping with the principle that the *trans* geometry for elimination is the "ideal" condition (see §12). All of the other stereoisomers possess at least

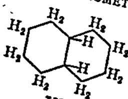


one pair of chlorine atoms *cis* to each other (thus having H and Cl *trans*). Cristol (1949) has also identified the α-isomer as the (±)-form.

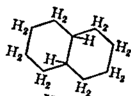
(vi) So far, we have discussed the stereochemistry of the *cyclo*hexane ring. The same types of stereoisomerism are also exhibited by various sized heterocyclic systems, *e.g.*, dimethyldiketopiperazine (§6. II), furanose (§7b. VII) and pyranose (§7a. VII) sugars.

(vii) *Decalins and decalols.* As we have seen, the boat and chair forms of *cyclo*hexane are readily interconvertible, and the result is that *cyclo*hexane behaves as if it were planar. Mohr (1918), however, elaborated Sachse's theory, and predicted that the fusion of two *cyclo*hexane rings, *e.g.*, as in decalin, should produce the *cis*- and *trans*-forms which would be sufficiently stable to retain their identities. This prediction has now been confirmed experimentally.

A non-committal way of writing the two geometrical isomers of decalin is given by formulae VII and VIII. On the other hand, several conventions have been introduced to represent these isomers. One convention uses *full* lines to represent groups *above* the plane

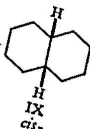


VII
cis-decalin

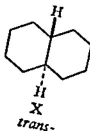


VIII
trans-decalin

of the molecule, and *broken* lines to represent those *below* the plane (*cf.* §5. xi); thus *cis*-decalin will be IX and *trans*-decalin X. This convention appears to be the one most widely used (see, *e.g.*, Steroids, Ch. XI), but there is another, introduced by Linstead (1937), which is favoured by many. According to this convention,



IX
cis-



X
trans-

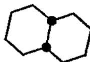
a hydrogen atom is represented as being above the plane of the ring when drawn as in XI, and below the plane when drawn as in XII; thus *cis*-decalin will be XIII, and *trans*-decalin XIV.



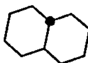
XI



XII

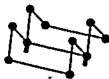


XIII
cis-



XIV
trans-

Fig. 7 shows the original diagrammatical method of representing *cis*-decalin by the fusion of two boat forms of *cyclo*hexane, and *trans*-decalin by the fusion of two chair forms; these are the forms suggested by Mohr.



cis-decalin



trans-decalin

Fig. 4.7.

The configurations of the decalins, however, are now known to be more complicated than this, the complication arising from the fact that a number of strainless modifications are possible, which differ in the type of "locking", *i.e.*, whether axial or equatorial bonds are used to fuse the rings. According to Hassel *et al.* (1946), *cis*- and *trans*-decalins are as shown in Fig. 8; the *cis*-form is produced by joining one axial and one equatorial bond of each ring, whereas

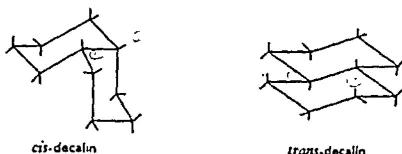
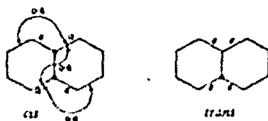


FIG. 4.8.

the *trans*-form is produced by joining the two rings by equatorial bonds only, in both cases the *cyclo*hexane rings are all chair forms (see also below).

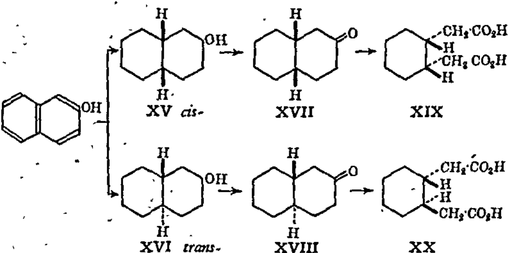
Johnson (1953) has calculated the difference in energy content between these two forms in the following simple manner. The *trans*-form is arbitrarily assigned a value of zero energy, and when this form is compared with the *cis*, it will be found that the latter has three extra skew interactions involving the two axial bonds (this is shown in the following diagram; the *cis*-form has 3 staggered and 15 skew arrangements, and the *trans*-form 6 staggered and 12 skew). Since each of these skew interactions is associated with



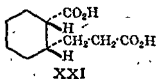
an energy increase of 0·8 kg cal., the total energy difference between the *cis*- and *trans*-forms is 3 × 0·8 = 2·4 kg cal. It might be noted, in passing, that if these two decalins are regarded as 1:2-disubstituted *cyclo*hexanes, then the *trans*-form (*ee*) would be expected to be more stable than the *cis*- (*ea*).

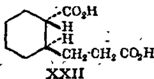
We shall now deal with the determination of configurations in the

decalin series. The configurations may be ascertained by using the Auwers-Skita rule (see §5. (x)*b*). Hückel (1923, 1925), however, isolated two forms of 2-decalol and determined their configurations by the following chemical methods. 2-Naphthol, on hydrogenation in the presence of nickel as catalyst, gave two 2-decalols, XV and XVI, each of which, on oxidation with chromic acid, gave a decal-2-one (XVII and XVIII). These two decalones each gave, on oxidation with permanganate, a *cyclo*hexane-1:2-diacetic acid. These diacetic acids were geometrical isomers; one was resolvable and therefore must be the *trans*-isomer, XX; and the other, which was not resolvable, must therefore be the *cis*-isomer, XIX (this is the *meso*-form). Thus the configurations of the two decalols and the two decalones are established:



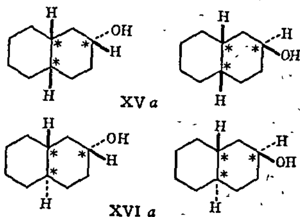
In addition to the two *cyclo*hexane-1:2-diacetic acids (which are formed by scission of the 2:3-bond of the decalone), two other geometrical isomers were also obtained, *viz.*, *cis*- and *trans*-*cyclo*hexane-1-carboxyl-2-propionic acids, XXI and XXII (these are formed by scission of the 1:2-bond of the decalone).





The conversion of 2-naphthol into two decalols does not prove that the two decalols are the *cis*- and *trans*-isomers described above. It is possible that both compounds could have been the *cis*- and *trans*-forms of a *given* decalol; since the carbon atom of the CHOH

group in the 2-decalol is asymmetric, it can exist in *two* configurations, *i.e.*, each decalol, XV and XVI, can exist in two forms;



XV*a*

XVI *a*

XV*a* and XVI*a*. Had the two decalols been the two forms of either XV or XVI, then on their oxidation, only *one* decalone would have been produced. Since, however, *two* decalones were obtained, the two decalols must be of the types XV and XVI—one of each, or even a mixture of the pairs; further proof of the existence of the types XV and XVI lies in the fact that the two decalones gave geometrical isomers of *cyclo*hexane-1:2-diacetic acid.

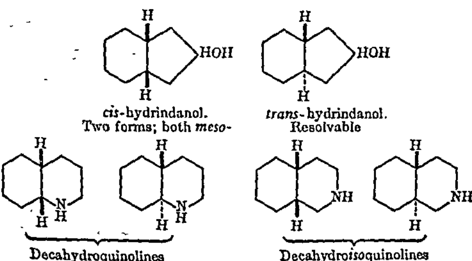
Consideration of formulae XV*a* and XVI*a* shows the presence of three asymmetric carbon atoms in each of the four possible forms, and since all four possess no elements of symmetry, four pairs of enantiomorphs should be possible theoretically. Actually all eight forms have been isolated, but their configurations have not yet been established with certainty.

There are only *two* geometrical isomers possible for the decalins, and their configurations have been established by the reduction of the two decalones, XVII and XVIII, by means of the Wolff-Kishner method (Eisenlohr *et al.*, 1924; see also Vol. I); each decalone gives the corresponding decalin. It is interesting to note in this connection that Willstatter *et al.* (1924) found that hydrogenation of naphthalene in the presence of platinum black as catalyst gives mainly *cis*-decalin, whereas in the presence of nickel as catalyst the main product is *trans*-decalin.

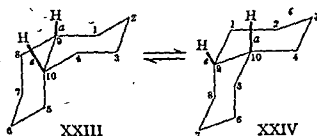
Various other fused ring systems have also been shown to exhibit the same type of geometrical isomerism as the decalins, *e.g.*, the hydrindanols exist in *cis*- and *trans*-forms (Hückel *et al.*, 1926), and also the decahydroquinolines and decahydro*iso*quinolines (Helfer *et al.*, 1923, 1926).

It has already been pointed out that in monosubstituted *cyclo*-

hexanes, the preferred conformation is the one with the substituent equatorial, but owing to the low energy barrier between this and the axial form, the two are readily interconvertible. In the case of the monosubstituted decalins, the problem is more complicated.

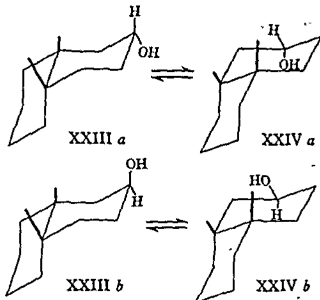


In *cis*-decalin, since ring fusion involves equatorial and axial bonds, the molecule is flexible and can interchange with the other *cis*-form, *i.e.*, there are two *cis*-forms possible (XXIII and XXIV), and these are identical and in equilibrium (*cf.* *cyclo*hexane). This has been shown to be so by Hassel (1950); thus:



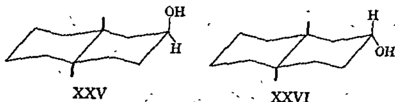
Now let us consider *cis*-2-decalol. Here there are four possible conformations which, in pairs, are in equilibrium. Two arise from XXIII (XXIII*a* and XXIII*b*), and two from XXIV (XXIV*a* and XXIV*b*).

In XXIII*a* and XXIV*b* the hydroxyl group is equatorial, and so these two conformations contain about the same energy. In XXIV*a* and XXIII*b* the hydroxyl group is axial, and on the basis that an equatorial conformation is more stable than an axial, then XXIII*a* and XXIV*b* will contribute more to the actual state of the molecule than will XXIV*a* and XXIII*b*, *i.e.*, the hydroxyl



group in *cis*-2-decalol should possess more equatorial character than axial. It is also interesting to note that the two axial *forms* do not contain the same energy. In XXIII*b* the *a*-hydroxyl group is involved in the normal 1:3-hydrogen interactions (at 4 and 9), but in XXIV*a* the interaction is the normal 1:3- with the hydrogen at 4 and the larger 1:3-interaction with the CH_3 group at 8. Thus XXIV*a* should be less stable than XXIII*b*.

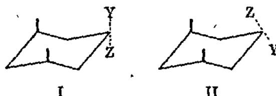
In *trans*-decalin there is only one stable conformation, since the ring fusions use equatorial bonds. If the molecular conformation were "inverted", the two ring fusions would now have to be axial, and this type of fusion is impossible (the axial bonds on adjacent carbon atoms are pointing in *opposite* directions). Thus, in *trans*-2-decalol, there are only two conformations possible, XXV and



XXVI. Furthermore, the latter, with the equatorial-hydroxyl conformation would be expected to be more stable than the former (with the axial hydroxyl).

§12. Effect of conformation on the course and rate of reactions. Since the environments of axial and equatorial groups are different, it may be expected that the reactivity of a given

group will depend on whether it is axial or equatorial. Now S_N2 reactions always occur with inversion (§4. III). Hence if the geometry of the molecule is such as to hinder the approach of the attacking group (Z) along the bonding line remote from the group to be expelled (Y), then the S_N2 reaction will be slowed down. Examination of formulae I and II shows that the transition state for an S_N2 reaction is more readily formed when Y is axial (I) than when it is equatorial (II). In I, the approach of Z is unhindered

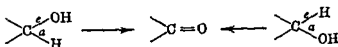


and the expulsion of Y assisted by the normal 1:3-interactions. In II, the approach of Z is hindered by the rest of the ring. Thus S_N2 reactions take place more readily with an axial substituent than with an equatorial. Using this as a generalisation, it is possible to assign conformation, *e.g.*, Fierens *et al.* (1952) have shown that *cyclo*hexyl bromide has a relatively low reactivity; the bromine is therefore equatorial in this compound (as would be expected in the absence of dipole interactions and hydrogen bonding).

The study of S_N1 reactions in *cyclo*hexane derivatives is made difficult because of the ease with which elimination reactions usually occur at the same time. It can be expected, however, that an S_N1 reaction will be sterically accelerated for an axial substituent, since the formation of a carbonium ion will relieve the steric strain due to 1:3-interactions. On the other hand, since these 1:3-interactions are absent for an equatorial substituent, no steric acceleration will operate in this conformation. In practice, however, it has been found that in S_N1 reactions where the substituent is restricted to an axial or to an equatorial position, the difference in reaction rate is much less than expected (from the above conformational arguments): *e.g.*, Winstein *et al.* (1955) have shown that the solvolysis of *cis*-4-*t*.-butyl*cyclo*hexyl-*a*-tosylate is only 3–4 times as fast as that of the corresponding *trans*-*e*-tosylate. The reasons for this behaviour are uncertain.

A particularly important substituent group in cyclic compounds is hydroxyl; and two very important reactions in which this group is involved are esterification and hydrolysis (of the ester). Owing to the hindered character of an axial group due to 1:3-interactions, esterification and hydrolysis will occur more readily with the

equatorial conformation. In the same way, esterification and hydrolysis of esters in which a carboxyl group is the substituent will also occur more readily when this group is equatorial. On the other hand, the relative rates of oxidation of secondary *a*- and *e*-alcohols to ketones by chromic acid (or hypobromous acid) is the reverse of the relative rates of hydrolysis of their carboxylic esters, *i.e.*, an *a*-hydroxyl is more readily oxidised than an *e*-. The reason for this is that the rate-determining step in this oxidation is a direct attack on the hydrogen atom of the C—H bond. If the hydroxyl is axial, the hydrogen is equatorial, and *vice versa*; thus:

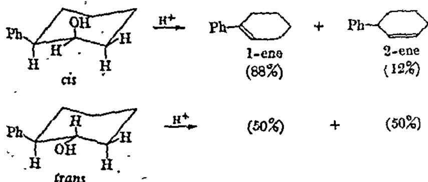


Elimination reactions are also of great importance in cyclic compounds. As we have seen (§5. XI), in ionic E2 reactions the four centres involved lie in a plane. In *cyclo*hexane systems this geometrical requirement is only found in *trans*-1:2-diaxial compounds, and these compounds thus undergo ready elimination reactions. In rigid systems, *e.g.*, the *trans*-decalin type, elimination in *trans*-1:2-diequatorial compounds is slower than in the corresponding diaxial compounds. *Cis*-1:2-Compounds (in which one substituent must be axial and the other equatorial) undergo elimination reactions slowly.

The steric course of E1 reactions is more difficult to study than that of E2 reactions because of the two-stage mechanism. This makes it difficult to ascertain the geometry of the intermediates involved. The formation of the carbonium ion will be sterically accelerated if the ionising group is axial and, if a second group is eliminated to form a double bond, this second stage will also be sterically accelerated if the second group is axial. Barton *et al.* (1951) have pointed out various examples in which E1 reactions are favoured by the diaxial conformation.

Not only does conformation control the rate of reactions, but it also may affect the course of a reaction. An example of the latter effect is the elimination reaction undergone by 2-phenyl*cyclo*hexanol in the presence of phosphoric acid to form phenyl*cyclo*hexene. Price *et al.* (1940) have shown that both the *cis* and *trans* alcohols are dehydrated, the former more readily than the latter. The product was shown to be a mixture of phenyl*cyclo*hex-1- and 2-ene, the former predominating when the *cis*-alcohol was used, and both olefins being present in about equal amounts when the *trans*-alcohol was used. These results may be explained as follows. In the *cis*-alcohol the conformations of the hydrogen and hydroxyl which

are eliminated are in the *trans* position (the "ideal" position), whereas in the *trans*-alcohol these two groups are in the *cis* position.



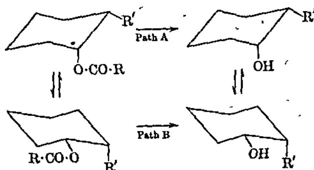
Another example of the effect of conformation on the course of a reaction in *cyclo*hexane systems is the action of nitrous acid on amines. Mills (1953) has proposed the following generalisation: When the amino-group is equatorial, the product is an alcohol with an equatorial conformation; but when the amino-group is axial, the main product is an olefin together with some equatorial alcohol.

Just as *trans* elimination is favoured with the two groups axial and *trans*, so it has been found that addition of electrophilic reagents to a double bond in *cyclo*hexenes is predominantly diaxial.

As we have seen, although there is a preferred form in *cyclo*hexane derivatives, the energy of interconversion between the preferred and less stable form is too low to permit their being distinguished by the classical methods of stereochemistry. This predominance of the preferred form holds good at room temperature (or below). At higher temperatures, or during the course of a chemical reaction, the preponderance of the preferred form may be reduced. In chemical reactions, it may be possible for the reaction to proceed more readily through the less stable conformation because it is this one which more closely approaches the geometry of the transition state. An example of this type is chloro*cyclo*hexane. As we have seen, the preferred form is the equatorial conformation. This compound, on treatment with ethanolic potassium hydroxide, undergoes dehydrohalogenation to form *cyclo*hexene. Since *trans* elimination is preferred, the reaction probably proceeds *via* the axial form.



Now let us consider reactions involving the hydroxyl group. It has already been pointed out that equatorial hydroxyl groups are more readily esterified, and equatorial esters more readily hydrolysed, than when these groups are axial. If an axial ester group has to stay in this position during hydrolysis, then because of the steric hindrance (1:3-interactions), the rate will be relatively slow (reaction path A). It is possible, however, that prior to reaction, the molecule is forced into the equatorial conformation (*cf.* chloro*cyclo*hexane above). If this were to happen, then the slower rate of hydrolysis would be due to the additional energy required to bring about the change in conformation (reaction path B).



Experimental data has enabled one path to be distinguished from the other (see also §16. VIII).

In *fused systems*, owing to the rigidity of the structure, such interconversions (as described above) are far less likely to occur.

In this chapter, the discussion of conformational analysis has been applied to *cyclo*hexane and its derivatives, and this has been done in order to introduce some of the ideas connected with this problem. The generalisations applicable to *cyclo*hexane compounds, however, are also applicable to heterocyclic compounds containing nitrogen, oxygen or sulphur (see, *e.g.*, tropines, §22. XIV; carbohydrates, §7h. VII). They are also applicable to the polynuclear compounds, *e.g.*, the Steroids; in fact, much of the work leading to these generalisations has been carried out on these compounds (see §4c. XI).

READING REFERENCES

Wheland, *Advanced Organic Chemistry*, Wiley (1949). Ch. 7 (p. 290). The Stereochemistry of Additions to Carbon-Carbon Double Bonds.

Ingold, *Structure and Mechanism in Organic Chemistry*, Bell and Sons (1953). Ch. 12. Additions and Their Retrogressions.

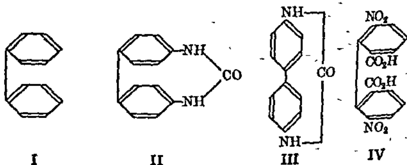
Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol. IV (1953). Ch. 12. Oxidation Processes.

de la Mare, Kinetics of Thermal Addition of Halogens to Olefinic Compounds, *Quart Reviews* (*Chem. Soc.*), 1949, **3**, 126.
Crombie, Geometrical Isomerism about Carbon-Carbon Double Bonds, *Quart. Reviews* (*Chem. Soc*), *1952*, **6**, *101*.
Hassel, Stereochemistry of *cyclo*Hexane, *Quart. Reviews* (*Chem. Soc.*), 1953, **7**, 221.
Bent, Aspects of Isomerism and Mesomerism, *J. Chem. Educ.*, 1953, **30**, 220, 284, 328.
Figueras, Stereochemistry of Simple Ring Systems, *J. Chem Educ.*, 1951, **28**, 134.
Winstein and Lucas, Loss of Optical Activity in the Reactions of the Optically Active *Erythro*- and *Threo*-3-bromo-2-butanols with Hydrobromic Acid, *J. Amer. Chem. Soc.*, 1939, **61**, 2845.
Klyne (Ed.), *Progress in Stereochemistry*, Butterworth (1954). Ch. 2. The Conformation of Six-membered Ring Systems.
Barton and Cookson, The Principles of Conformational Analysis, *Quart. Reviews* (*Chem Soc.*), 1956, **10**, 44.
Orloff, The Stereoisomerism of *cyclo*Hexane Derivatives, *Chem. Reviews*, 1954, **54**, 347.
Newman (Ed.), *Steric Effects in Organic Chemistry*, Wiley (1956) Ch. 1. Conformational Analysis.
Wyman, The *Cis-Trans* Isomerisation of Conjugated Compounds, *Chem. Reviews*, 1955, **55**, 625.
Wicker, The Mechanism of Catalytic Hydrogenation of Cyclic Compounds, *J.C.S*, **1956**, 2165.

CHAPTER V

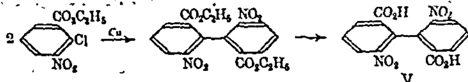
STEREOCHEMISTRY OF DIPHENYL COMPOUNDS

§1. Configuration of the diphenyl molecule. If we assume that the benzene ring is planar, then the diphenyl molecule will consist of two planar rings; but without any further information we cannot say how these two rings are arranged spatially. Kaufler (1907) proposed the "butterfly" formula, I, in order to account for the chemical behaviour of various diphenyl derivatives, *e.g.*, Michler and Zimmermann (1881) had condensed benzidine with



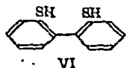
carbonyl chloride and obtained a product to which Kaufler assigned structure II. According to Kaufler, the co-axial structure III was impossible, since the two amino-groups are too far apart to react simultaneously with carbonyl chloride; it should be noted that this *simultaneous* reaction at both ends was *assumed* by Kaufler. Simultaneous reaction, however, is reasonable (according to Kaufler) on the folded structure, II.

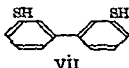
Now Schultz (1880) had prepared a dinitrodiphenic acid by the nitration of diphenic acid, and Schmidt *et al.* (1903), from their work on this acid, believed it to be 6.6′-dinitrodiphenic acid, IV; these workers, it should be noted, did not synthesise the acid. In 1921, however, Kenner *et al.* synthesised 6.6′-dinitrodiphenic acid by means of the Ullmann reaction (see Vol. I) on the ethyl ester of 2-chloro-3-nitrobenzoic acid, and hydrolysing the product. This acid, V (written with the two benzene rings co-axial), did not have the same melting point as Schultz's acid, and so Kenner, believing that his and Schultz's acid were both 6.6′-dinitrodiphenic acid, suggested that the two were stereoisomers. Then Christie and Kenner (1922) showed that Kenner's acid was resolvable, and

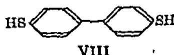


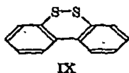
pointed out that this could be explained on the Kaufler formula, IV, since this structure has no elements of symmetry. These authors, however, also pointed out that the optical activity could also be accounted for by the co-axial structure, V, provided that the two benzene rings do not lie on one plane (see also §2).

Kaufler's formula, as we have seen, was based on the assumption that the two amino-groups in benzidine react *simultaneously* with various reagents. Re-investigation of these reactions showed that this was not the case, *e.g.*, Turner and Le Fèvre (1926) found that the compound produced from benzidine and carbonyl chloride was not as originally formulated (see II or III), but had a free amino-group, *i.e.*, the compound was $[NH_2{\cdot}C_6H_4{\cdot}C_6H_4{\cdot}NH]_2CO$. Hence Kaufler's *reason* for his butterfly formula is incorrect, and although it does not necessarily follow that the *formula* is incorrect, nevertheless Turner's work weakened Kaufler's claim. One of the strongest bits of chemical evidence for rejecting Kaufler's formula is that of Barber and Smiles (1928). These workers prepared the three dimercaptodiphenyls, VI, VII and VIII, and oxidised each









one. Only one of them, the 2:2'-derivative, VI, gave the intramolecular disulphide (diphenylene disulphide, IX). On the Kaufler formula, all three dithiols would be expected to give the intramolecular disulphides, since the two thiol groups are equally distant in all three compounds.

Physico-chemical methods have also been used to determine the configuration of the diphenyl molecule, *e.g.*, the crystal structure of 4:4'-diphenyl derivatives shows a centre of symmetry; this is only possible for the co-axial formula. Dipole moment measurements also confirm this configuration, *e.g.*, the dipole moment of

4:4'-dichlorodiphenyl is zero; this again is only possible if the two benzene rings are co-axial.

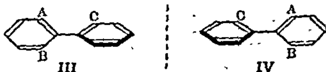
§2. **Optical activity of diphenyl compounds.** Christie and Kenner's work (see above) has been extended by other workers, who showed that compounds in which at least *three* of the four *ortho*-positions in diphenyl are occupied by certain groups could be resolved. It was then soon found that two conditions were necessary for diphenyl compounds to exhibit optical activity:

(i) Neither ring must have a vertical plane of symmetry. Thus I is not resolvable, but II is.



(ii) The substituents in the *ortho*-positions must have a large size, *e.g.*, the following compounds were resolved: 6-nitrodiphenic acid, 6.6'-dinitrodiphenic acid, 6:6'-dichlorodiphenic acid, 2.2'-diamino-6.6'-dimethyldiphenyl (see also §4).

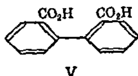
The earlier work showed that three groups had to be present in the *ortho*-positions. This gave rise to the theory that the groups in these positions impinged on one another when free rotation was attempted, *i.e.*, the spatial effect prevented free rotation. This theory of restricted rotation about the single bond joining the two benzene rings (in the co-axial formula) was suggested simultaneously in 1926 by Turner and Le Fèvre, Bell and Kenyon, and Mills. Consider molecule III and its mirror image IV. Provided that the groups A, B and C are large enough to "interfere mechanically", *i.e.*, to behave as "obstacles", then free rotation about the single bond is restricted. Thus the two benzene rings cannot be coplanar

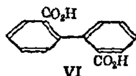


and consequently IV is not superimposable on III, *i.e.*, III and IV are enantiomorphs. In molecule III there is no asymmetric carbon atom; it is the molecule *as a whole* which is asymmetric, due to the restricted rotation.

In diphenyl the two benzene rings are co-axial, and in optically

active diphenyl derivatives the rings are inclined to each other due to the spatial and repulsive effects of the groups in the *ortho*-positions. The actual angle of inclination of the two rings depends on the nature of the substituent groups, but it appears to be usually in the vicinity of 90°, *i.e.*, the rings tend to be approximately perpendicular to each other. Thus, in order to exhibit optical activity, the substituent groups in the *ortho*-positions must be large enough

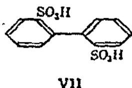


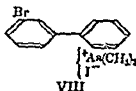


to prevent the two rings from becoming coplanar, in which case the molecule would possess a plane or a centre of symmetry, *e g.*, diphenic acid is not optically active. In configuration V the molecule has a plane of symmetry, and in configuration VI a centre of symmetry; of these two, VI is the more likely because of the repulsion between the two carboxyl groups (*cf.* §4. II).

If restricted rotation in diphenyl compounds is due entirely to the spatial effect, then theoretically we have only to calculate the size of the group in order to ascertain whether the groups will impinge and thereby give rise to optical activity. In practice, however, it is found that groups (and atoms) behave as if they were larger than the volumes obtained from group (and atomic) radii (*cf.* §15b. I). This behaviour is largely due to the fact that groups also repel (or attract) one another because of the electric charges that are usually present on these groups. Thus the actual distance that the atoms or groups (in the *ortho*-positions) can approach one another is *greater* than that obtained from the atomic and group radii. Better agreement with experiment is obtained when the van der Waals radii (§2. I) are used for calculating the "size" of a group.

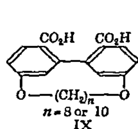
Later work has shown that if the substituent groups are large enough, then only *two* in the *o*- and *o'*-positions will produce restricted rotation, *e.g.*, Lesslie and Turner (1932) resolved diphenyl-2:2'-disulphonic acid, VII. In this molecule the sulphonic acid group is large enough to be impeded by the *ortho*-hydrogen atoms.



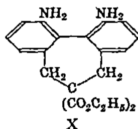


Lesslie and Turner (1933) have also resolved the arsonium compound VIII; here also the trimethylarsonium group is large enough to be impeded by the *ortho*-hydrogen atoms (the bromine atom in the *meta*-position gives asymmetry to this ring). This example is unique up to the present in that only *one* substituent in the *ortho*-position produces optical activity in diphenyl compounds.

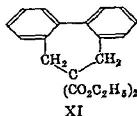
It has already been pointed out that diphenic acid is not optically active, and that its configuration is most probably VI. Now calculation shows that the effective diameter of the carboxyl group is large enough to prevent configuration V from being planar, and consequently, if the two rings could be held more or less in this configuration, the molecule would not be coplanar and hence would



IX



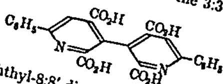
X



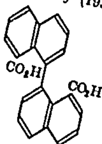
XI

be resolvable. Such a compound, IX, was prepared and resolved by Adams and Kornblum (1941). The two benzene rings are not coplanar and are held fairly rigid by the large methylene ring. Iffland *et al.* (1956) have also prepared the optically active diphenyl X which has a 2·2′-bridge and two amino-groups in the 6.6′-positions. On the other hand, these authors have also prepared XI in optically active forms; this compound has the 2:2′-bridge but no substituents in the 6 6′-positions.

Adams *et al.* (1932) have also resolved the 3:3′-dipyridyl

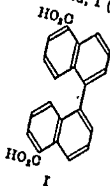


(ii) 1:1′-Dinaphthyl-8:8′-dicarboxylic acid has been obtained in optically active forms by Stanley (1931).

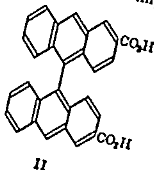


This compound gives rise to asymmetric transformation (§10 iv. II); resolution with brucine gave 100 per cent. of either the (+)- or (−)-compound.

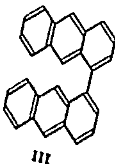
Other compounds similar to the dinaphthyl which have been obtained in optically active forms are 1:1′-dinaphthyl-5:5′-dicarboxylic acid, I (Bell *et al.*, 1951), the dianthryl derivatives, II and



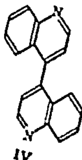
I



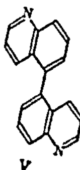
II



III



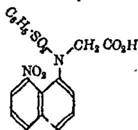
IV



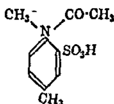
V

III (Bell *et al.*, 1949), and the 4:4′- and 5:5′-diquinolyls, IV and V (Crawford *et al.*, 1952).

(iii) Mills and Elliott (1928) obtained *N*-benzenesulphonyl-8-nitro-1-naphthylglycine, VI, in optically active forms; these were optically unstable, undergoing asymmetric transformation with brucine. Mills and Kelham (1937) also resolved *N*-acetyl-*N*-methyl-*p*-toluidine-3-sulphonic acid, VII, with brucine, and found

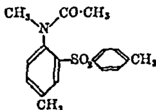


VI

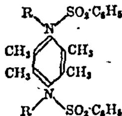


VII

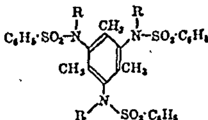
that it racemised slowly on standing. In both VI and VII the optical activity arises from the restricted rotation about the C—N bond (the C being the ring carbon to which the N is attached). Asymmetry arising from the same cause is also shown by 2-acetomethylamido-4′:5-dimethyldiphenylsulphone, VIII; this was partially resolved by Buchanan *et al.* (1950; see also §10 iv. II). It is also interesting to note in this connection that Adams *et al.* (1950) have isolated pairs of *geometrical* isomers of compounds of the types IX and X; here geometrical isomerism is possible because of the restricted rotation about the C—N bonds.



VIII

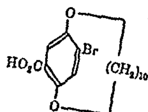


IX



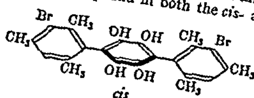
X

(iv) Lüttringhaus *et al.* (1940, 1947) isolated two optically active forms of 4-bromogentisic acid decamethylene ether. In this compound the methylene ring is perpendicular to the plane of

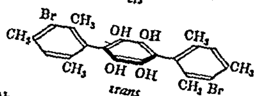


the benzene ring; the two substituents, Br and CO_2H, prevent the rotation of the benzene nucleus inside the large ring.

(v) Terphenyl compounds can exhibit both geometrical and optical isomerism when suitable substituents are present to prevent free rotation about single bonds, *e.g.*, Shildneck and Adams (1931) obtained the following compound in both the *cis*- and *trans*-forms.

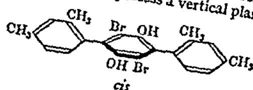


cis

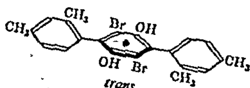


trans

Interference of the methyl and hydroxyl groups in the *ortho*-positions prevents free rotation and tends to hold the two outside rings perpendicular to the centre ring. Inspection of these formulae shows that if the centre ring does not possess a vertical plane of symmetry,



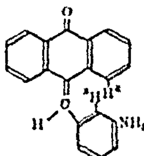
cis



trans

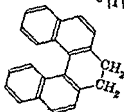
then optical activity is possible. Thus Browning and Adams (1930) prepared the following *cis*- and *trans*-forms, and resolved the *cis*-isomer; the *trans*-isomer is not resolvable since it has a centre of symmetry.

(vi) A very interesting case of restricted rotation about a single bond is afforded by the compound 10-*m*-aminobenzylideneanthrone. This was prepared by Ingram (1950), but he failed to resolve it.

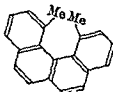


He did show, however, that it was optically active by the mutarotation of its camphorsulphonate salt, and by the preparation of an active hydrochloride. Thus the molecule is asymmetric, and this asymmetry can only be due to the restricted rotation of the phenyl group about the C—phenyl bond, the restriction being brought about by hydrogen atoms in the *ortho*-positions. The two hydrogen atoms labelled H* overlap in space, and consequently the benzene ring cannot lie in the same plane as the 10-methyleneanthrone nucleus.

about 3·0 A. Thus, if the geometry of the molecule is such as to produce "intramolecular overcrowding", the molecule becomes distorted. An example of this type is 4:5:8-trimethyl-1-phenanthrylacetic acid, I. The phenanthrene nucleus is planar and substituents lie in this plane. If, however, there are fairly large groups in positions 4 and 5, then there will not be enough room to accommodate both groups in the plane of the nucleus. This leads to strain being produced by intramolecular overcrowding, and the strain may be relieved by the bending of the substituents out of the plane of the nucleus, or by the bending (buckling) of the aromatic rings, or by both. Whichever explanation is correct, the molecule will no longer be planar owing to out-of-plane distortions. If the molecule is not planar, it will be asymmetric and hence it should be (theoretically) resolvable. Newman *et al.* (1940, 1947) have actually partially resolved it, and have also partially resolved II and III (both of which also exhibit out-of-plane distortions). All of these compounds were found to have low optical stability, but Turner *et al.* (1955) have prepared the optically active forms of 9:10-dihydro-3:4-5:6-dibenzophenanthrene (IV), which is more optically stable than I, II and III.



IV



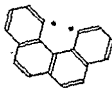
V



VI

Newman *et al.* (1955, 1956) have prepared V and VI which, so far, are the most optically stable compounds of the intramolecular overcrowding type.

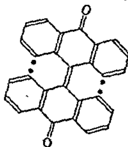
It will be noticed that in IV and VI the only way in which out-of-plane distortion can occur is through buckling of the molecule.



VII



VIII



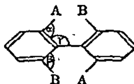
IX

The simplest molecule exhibiting overcrowding and consequent *out-of-plane buckling* of the molecule is 3:4-benzophenanthrene (VII); this has been shown to be non-planar by X-ray analysis (Schmidt *et al.*, 1954). Similarly, Robertson *et al.* (1954) have shown that VIII exhibits out-of-plane buckling.

Another point to note in connection with out-of-plane buckling is that the buckling is distributed over all the rings in such a manner as to cause the minimum distortion in any one ring. This distortion, which enables non-bonded carbon atoms to avoid being closer together than 3·0 A (marked with dots in VII and VIII), forces some of the other carbon atoms to adopt an almost tetrahedral valency arrangement (the original hybridisation is trigonal), and this affects the physical and chemical properties of the molecule, *e.g.*, Coulson *et al.* (1955) have calculated that the deformation in VIII produces a loss of resonance energy of about 18 kg.cal./mole.

Just as benzene rings may suffer distortion, so can a molecule which owes its planarity to the presence of a double bond. Such an example is dianthronylidene (IX). The carbon atoms marked with dots are overcrowded (the distance between each pair is 2·9 A), and the strain is relieved by a rotation of about 40° around the olefinic double bond (Schmidt *et al.*, 1954).

§4. Racemisation of diphenyl compounds. Since the optical activity of diphenyl compounds arises from restricted rotation, it might be expected that racemisation of these compounds would not be possible. In practice, it has been found that many optically active diphenyl compounds can be racemised under suitable conditions, *e.g.*, boiling in solution. The general theory of these racemisations is that heating increases the amplitude of the vibrations of the substituent groups in the 2:2′:6:6′-positions, and also the amplitude of vibration of the two benzene rings with respect to each other, thereby permitting the substituent groups to slip by one another. Thus the nuclei pass through a common plane and hence the probability is that the final product will contain an



equimolecular amount of the (+)- and (−)-forms. Westheimer (1946–1950) has assumed, in addition to the above bond-stretchings, that the angles α, β and γ are deformed, and also the benzene rings themselves are deformed during racemisation.

The foregoing theory of racemisation is analogous to Werner's theory for the racemisation of compounds which contain an asymmetric carbon atom. According to Werner (1904), the groups in the compound *Cabde* are set vibrating under the influence of heat, and if the amplitude of vibration becomes large enough, all four groups will become coplanar at some instant (Fig. 1). This planar

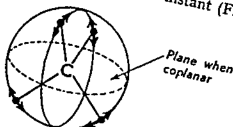
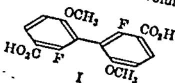


Fig. 5.1.

structure is symmetrical, and when the molecule emerges from this condition, there is an equal chance of its doing so in the (+)- or (−)-configuration, *i.e.*, the molecule racemises. There is, however, a great deal of evidence against this mechanism in compounds of the type *Cabde*, *e.g.*, from spectroscopic data it appears that the bonds would break before the vibrations were large enough to permit a planar configuration to be reached. Furthermore, Kincaid and Henriques (1940), on the basis of calculations of the energy required for the inversion of molecules, were led to suggest that the molecule *Cabde* can only be racemised by the bonds actually breaking. Even so, this theory of racemisation appears to be the most reasonable one for the racemisation of diphenyl compounds. In this case, the amplitude of vibration does not have to be large in order to permit the *ortho*-groups to slip by one another. This is supported by the fact that it has been found that diphenyl compounds with small substituent groups racemise easily, whereas when the groups are large, racemisation is difficult or even impossible.

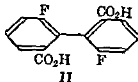
2:2′:6:6′-Tetrasubstituted diphenyl compounds may be classified under three headings according to the nature of the substituent groups.

(i) *Non-resolvable.* These contain any of the following groups: hydrogen, methoxyl or fluorine. The volumes (*effective* volumes)

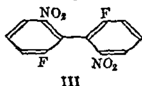


of these groups are too small to prevent rotation about the single bond. Thus 2:2'-difluoro-6 6'-dimethoxydiphenyl-3:3'-dicarboxylic acid, I, is non-resolvable.

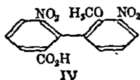
(ii) *Resolvable, but easily racemised.* These must contain at least two amino-groups, or two carboxyl groups, or one amino- and one carboxyl group; the remaining groups may be any of those given in (i) [but not hydrogen]. Thus 6.6'-difluorodiphenic acid, II, is resolvable, and is readily racemised.

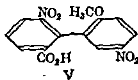


(iii) *Not racemisable at all.* Diphenyl compounds which fall in this group are those which contain at least two nitro-groups; the other groups can be any of those given in (i)—but not hydrogen—and (ii). Thus 2:2'-difluoro-6:6'-dinitrodiphenyl, III, is resolvable, and cannot be racemised.

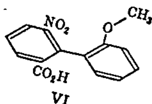


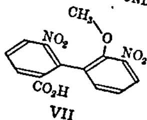
In addition to the size of the groups in the *ortho*-positions, the nature and position of other substituent groups also play a part in the rate of racemisation, *e.g.*, the rate of racemisation of IV is much slower than that of V (Adams *et al.*, 1932, 1934). Thus the





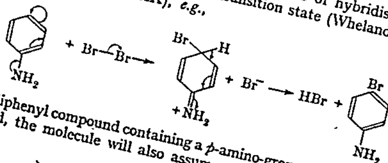
nitro-group in position 3' has a much greater stabilising influence than in position 5'. The reason for this is uncertain, but one possible explanation is as follows. In VI, the methyl group of the methoxyl group is probably in the configuration shown. In VII, the nitro-group in the 3'-position would tend to force the methyl group away, the resulting configuration being somewhat as shown in VII; in



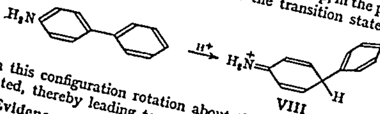


this condition there would be greater interference between the methoxyl group and the two groups in the other benzene ring.

If racemisation of diphenyl compounds is purely a *physical* phenomenon (as has been described above), then one would not expect the velocity of racemisation to be affected by catalysts. Crawford *et al.* (1954), however, have suggested that the racemisation would be facilitated by the presence of activating groups, *e.g.*, the amino-group, especially in the *para*-positions. When the benzene ring is attacked by electrophilic reagents, the attacked carbon atom changes from the trigonal state of hybridisation to the tetrahedral state to form the transition state (Wheland, 1942; see also Vol. I, Ch. XX), *e.g.*,



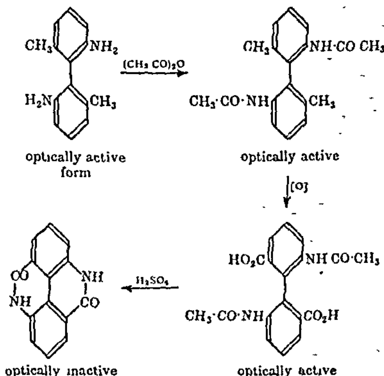
In a diphenyl compound containing a *p*-amino-group, in the presence of acid, the molecule will also assume the transition state, VIII,



and in this configuration rotation about the interannular bond is facilitated, thereby leading to racemisation.

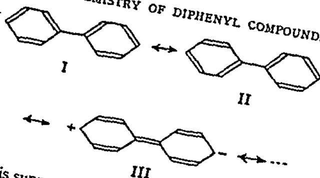
§5. Evidence for the obstacle theory. Evidence for the obstacle theory, *i.e.*, interference of groups, amounts to proving that the two benzene rings in optically active diphenyl compounds are not coplanar. A direct chemical proof for the non-coplanar configuration was given by Meisenheimer *et al.* (1927). The method was to unite the "obstacle groups" in optically active diphenyl compounds, thereby forming five- or six-membered rings. Now

such systems are known to be planar, and hence optical activity should disappear; this was found to be so in practice. Meisenheimer started with 2:2'-diamino-6.6'-dimethyldiphenyl, resolved it and then carried out the following reactions on one of the enantiomorphs:

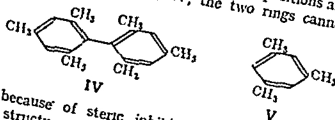


In all the optically active compounds, the rings *cannot* be coplanar, since if they were, the molecules would possess a centre or plane of symmetry. If the dilactam, however, is *not* planar, then it would possess no elements of symmetry, and consequently would be optically active. If the dilactam is planar, then it has a centre of symmetry, and consequently cannot be optically active. This compound was, in fact, not optically active, and so must be planar.

Further evidence for the non-coplanar configuration of optically active diphenyl compounds has been obtained from physico-chemical studies. The ultraviolet absorption spectrum of diphenyl in the crystalline state is different from that of benzene. This difference can be explained on the basis that diphenyl is a resonance hybrid, some of the contributing resonating structures of which are not possible in benzene itself, *e.g.*, III. Thus the interannular bond will possess some partial double bond character, and consequently the molecule will be planar. The presence of this partial



double bond is supported by the fact that the distance between the two rings is 1·48 A; the C—C bond is 1·54 A, and the C=C bond 1·33 A (see Vol. I, Ch. XX). When the *ortho*-positions are occupied as in, for example, dimesityl, IV, the two rings cannot become

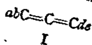


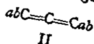
coplanar because of steric inhibition of resonance. Thus the resonating structures of IV will be the same as those of mesitylene, V, both of these being similar to the resonating structures of benzene. Pickett *et al.* (1936) have shown that the ultraviolet absorption spectra of IV and V are the same, and totally different from the ultraviolet absorption spectrum of diphenyl.

X-ray analysis and ultraviolet absorption measurements show that the diphenyl molecule is planar in the *crystalline* state. Electron diffraction studies of the *vapour* show that the two rings are inclined to each other at approximately 45°. Also, ultraviolet absorption and dipole measurements of diphenyl in *solution* show that the two rings are not coplanar. It has been suggested that in any physical state other than the crystalline, the steric repulsion between the 2:2′-hydrogen atoms tends to make the molecule non-planar, in opposition to the planar configuration due to resonance.

§6. STEREOCHEMISTRY OF THE ALLENES

Allenes are compounds which have the general structure I.





Examination of the space formula of compounds of this type shows

that the molecule and its mirror image are not superimposable. The modern way of writing I is shown in Fig. 2. The two end carbon atoms are in a state of trigonal hybridisation, and the centre carbon atom is in the digonal state. Thus the centre carbon atom forms two π-bonds which are perpendicular to each other; in Fig. 2 the π_x-bond is perpendicular to the plane of the

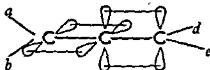
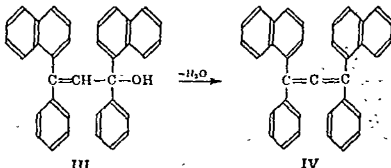


FIG. 5 2.

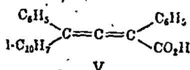
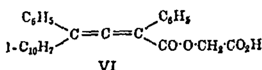
paper, and the π_y-bond is in the plane of the paper. In the trigonal state, the π-bond is perpendicular to the plane containing the three σ-bonds (see Vol. I, Ch. II); consequently the groups *a* and *b* lie in the plane of the paper, and the groups *d* and *e* in the plane perpendicular to the plane of the paper. This molecule does not possess a plane or centre of symmetry; this is also true for molecule II. Thus I and II will be resolvable.

The resolvability of allenes was predicted by van't Hoff in 1875, but experimental verification was not obtained until 1935, when Mills and Maitland carried out a catalytic asymmetric dehydration on $\alpha{:}\gamma$-di-1-naphthyl-$\alpha{:}\gamma$-diphenylallyl alcohol, III, to give the



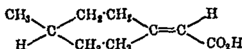
dinaphthyldiphenylallene, IV. When the dehydration was carried out with an optically inactive dehydrating catalyst, *e.g.*, *p*-toluenesulphonic acid, the racemic modification of the allene derivative was obtained. When, however, the alcohol III was boiled with a one per cent. benzene solution of (+)-camphorsulphonic acid, a dextrorotatory allene was obtained. Similarly, (—)-camphorsulphonic acid gave a lævorotatory allene.

The first successful *resolution* of an allene derivative was carried out by Kohler *et al.*, also in 1935. Lapworth and Wechsler (1910)

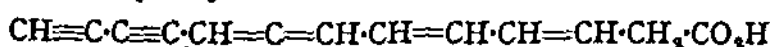
prepared γ-1-naphthyl-α:γ-diphenylallene-α-carboxylic acid, V, but failed to resolve it; they were unable to crystallise the salts with active bases. Kohler converted this acid into the glycollic acid ester, VI, and was then able to resolve VI by means of brucine.

Although allenes were not successfully resolved until 1935, compounds with a *similar* configuration were resolved as early as 1909. In this year, Pope *et al.* resolved 1-methyl*cyclo*hexylidene-4-acetic



acid, VII; in this compound one of the double bonds of allene has been replaced by a six-membered ring, and the general shape of the allene molecule is retained.

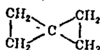
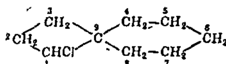
It is interesting to note, in connection with allenes, that the antibiotic *mycomycin* has been shown to contain the allene grouping. Mycomycin is optically active, and is the only known natural compound which owes its optical activity to the presence of this grouping. Celmer and Solomons (1953) have shown that the structure of mycomycin is:



$$CH{\equiv}C{\cdot}C{\equiv}C{\cdot}CH{=}C{=}CH{\cdot}CH{=}CH{\cdot}CH{=}CH{\cdot}CH_2{\cdot}CO_2H$$

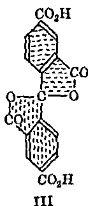
§7. STEREOCHEMISTRY OF THE SPIRANS

If both double bonds in allene are replaced by ring systems, the resulting molecules are *spirans*. One method of naming spirans obtains the root name from the number of carbon atoms in the *nucleus*; this is then prefixed by the term "spiro", and followed by numbers placed in square brackets which indicate the number of carbon atoms joined to the "junction" carbon atom. The positions of substituents are indicated by numbers, the numbering

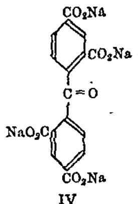



beginning with the *smaller* ring and ending on the junction carbon atom; *e.g.*, I is spiro-[2:2]-pentane, II is 1-chlorospiro-[5:3]-nonane.

Examination of these formulae shows that the two rings are perpendicular to each other, and hence suitable substitution will produce molecules with no elements of symmetry, thereby giving rise to optically active forms, *e.g.*, Mills and Nodder (1920, 1921) resolved the dilactone of benzophenone-2:2':4:4'-tetracarboxylic



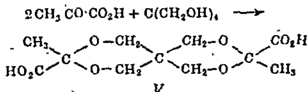
III



IV

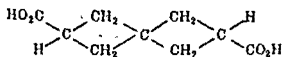
acid, III. In this molecule the two shaded portions are perpendicular to each other, and consequently there are no elements of symmetry. When this compound is treated with sodium hydroxide, the lactone rings are opened to form IV, and the optical rotation disappears.

Böeseken *et al.* (1928) condensed penta-erythritol with pyruvic acid and obtained the spiro-compound V, which they resolved. Some other spiro-compounds that have been resolved are the spiro-

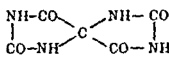


heptane, VI (Backer *et al.*, 1928, 1929), the spiro-hydantoin, VII (Pope and Whitworth, 1931), and the spiroheptane, VIII (Jansen and Pope, 1932).

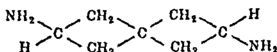
In all the cases so far discussed, the optical activity of the spiran is due to the asymmetry of the molecule as a whole; thus there is only one pair of enantiomorphs. If a spiro-compo[illegible] contains asymmetric carbon atoms, then the number of op[illegible] forms



VI

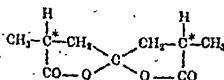


VII

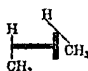


VIII

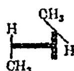
spiro-compound IX, which contains two similar asymmetric carbon atoms (*). If we imagine the left-hand ring of IX to be horizontal, then the right-hand ring will be vertical; and if we represent them



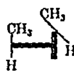
IX



X



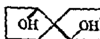
XI



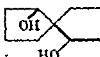
XII

by bold horizontal and vertical lines, respectively, then there are three different geometrical isomers possible, X, XI and XII (this can be readily demonstrated by means of models). Each of these geometrical isomers has no elements of symmetry, and so each can exist as a pair of enantiomorphs. Three racemic modifications were actually isolated by Sutter and Wijkman, but were not resolved.

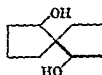
Cram *et al.* (1954) have also prepared the following three spiro [4:4] nonanediols (as racemates):



cis-cis



cis-trans



trans-trans

Various spiro-compounds have been prepared in which the spiro-atom is nitrogen (§2a. VI), phosphorus (§3b. VI), or arsenic (§4a. VI).

Reading References

Stewart and Graham, *Recent Advances in Organic Chemistry*, Longmans, Green. Vol. III (1948, 7th ed.). Ch. 11. The Diphenyl Problem.

Adams and Yuan, The Stereochemistry of Diphenyls and Analogous Compounds, *Chem. Reviews*, 1933, **12**, 261.

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). Vol. I. Ch. 4, pp. 337-382.

Crawford and Smyth, The Effect of Groups in Non-Blocking Positions on the Rate of Racemisation of Optically Active Diphenyls, *Chem. and Ind.*, **1954**, 346.

Ann. Reports (*Chem. Soc.*), Stereochemistry of Diphenyl Compounds, 1926, **23**, 119; 1931, **28**, 394; 1932, **29**, 69; 1935, **32**, 246; 1939, **36**, 255; 1953, **50**, 154; 1955, **52**, 131.

Klyne and de la Mare (Ed.), *Progress in Stereochemistry*, Butterworth. Vol. II (1958). Ch. I, p. 22. Molecular Overcrowding.

CHAPTER VI

STEREOCHEMISTRY OF SOME ELEMENTS OTHER THAN CARBON

§1. Hybridisation of orbitals. Many elements other than carbon form compounds which exhibit optical isomerism. Since the criterion for optical activity must be satisfied, *viz.*, the molecule must not be superimposable on its mirror image, it therefore follows that the configurations of the various molecules can never be planar. A given " central " atom may have different configurations according to the valency that atom exhibits ; the following table shows a number of common hybridisations.

Number of bonds	Orbitals used	Geometrical arrangement	Examples
2	*sp* or *dp*	Linear	—C≡C—, $HgCl_2$
2	p^2	Angular	H_2O, H_2S
3	sp^2	Trigonal plane	>C=C<, BCl_3
3	p^3	Trigonal pyramid	NH_3, PH_3, AsH_3
4	sp^3	Tetrahedral	CH_4, $(NH_4)^+Cl^-$, $SnCl_4$
4	dsp^2	Square planar	$Ni(CN)_4^=$, $PtCl_4^=$
5	dsp^3	Trigonal bipyramid	PCl_5
6	d^2sp^3	Octahedral	SF_6, $PtCl_6^=$

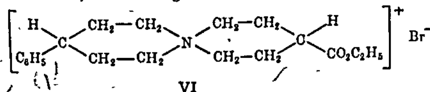
A knowledge of the shape of a given molecule may be used to deduce the nature of the orbitals that have been used in the formation of the bonds in that molecule. On the other hand, it has been possible to deduce the shapes of simple molecules from a knowledge of the orbitals that are used to form the bonds.

§2. STEREOCHEMISTRY OF NITROGEN COMPOUNDS

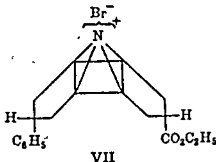
According to the electronic theory of valency, nitrogen can be tercovalent or quadricovalent unielectrovalent ; in both of these states nitrogen, as the " central " atom, can exhibit optical activity.

§2a. Quaternary ammonium salts. Originally, the valency of nitrogen in quaternary ammonium salts was believed to be quinquevalent ; later, however, it was shown that one valency was different from the other four. Thus, using the formula, $[Nabcd]^+X^-$, for quaternary ammonium salts, and assuming that

Werner (1911) had suggested the tetrahedral configuration, and this was supported by Neagi (1919) and Mills and Warren (1925). It was, however, Mills and Warren who gave the most conclusive evidence that the configuration is tetrahedral. Their evidence is based on the following argument. Compounds of the type *ab*C=C=C*ab* are resolvable since carbon is "tetrahedral" (see allenes, §6. V), and if nitrogen is also "tetrahedral", then the compound *ab*C=N=C*ab* should be resolvable, but will not be resolvable if the nitrogen is pyramidal. Mills and Warren prepared 4-carbethoxy-4′-phenylbispiperidinium-1:1′-spiran bromide, and resolved it. If the configuration of this molecule is VI, *i.e.*, a spiran, then it possesses no elements of symmetry, and hence will be resolvable; if the configuration is VII (*i.e.*, pyramidal), then it

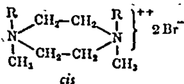


VI

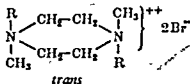


VII

will possess a vertical plane of symmetry, and hence will be optically inactive. Since the compound was resolved, the configuration must be tetrahedral, *i.e.*, VI. This tetrahedral configuration has been confirmed by physico-chemical studies (see §2b). More recently, Hanby and Rydon (1945) have shown that the diquaternary salts of dimethylpiperazine exhibit geometrical isomerism, and this is readily explained on the tetrahedral configuration of the four nitrogen valencies (*cf.* *cyclo*hexane, §11. IV).

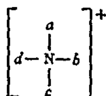


cis

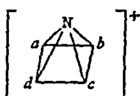


trans

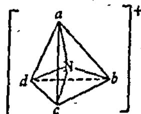
the charge on the nitrogen atom has no effect on the configuration of the cation, the cation may be considered as a five-point system similar to that of carbon in compounds of the type *Cabde*. This similarity is based on the assumption that the four valencies in the ammonium ion are equivalent, and this assumption is well substantiated experimentally and also theoretically. Hence there are three possible configurations for the cation $[Nabcd]^+$, I, II



I

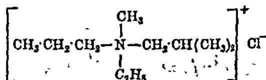


II



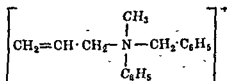
III

and III (*cf.* §3a. II). If the cation is planar (I), then it would not be resolvable; it would be resolvable, however, if the configuration is pyramidal (II) or tetrahedral (III). Le Bel (1891) claimed to have partially resolved *iso*butylethylmethylpropylammonium chloride, IV, by means of *Penicillium glaucum* (*cf.* §10 iii. II), but



IV

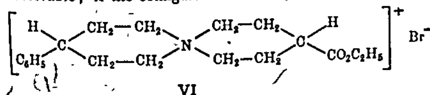
later work apparently showed this was wrong. The first definite resolution of a quaternary ammonium salt was that of Pope and Peachey (1899), who resolved allylbenzylmethylphenylammonium iodide, V, by means of (+)-bromocamphorsulphonic acid. This



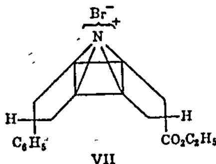
V

was the first case of optical activity due to a "central" atom other than carbon. This resolution was then followed by the work of Jones (1905), who resolved benzylethylmethylphenylammonium iodide. Thus the ammonium ion cannot be planar, but must be either pyramidal or tetrahedral. Bischoff (1890) had proposed a pyramidal structure, and this configuration was supported by Jones (1905) and Jones and Dunlop (1912). On the other hand,

Werner (1911) had suggested the tetrahedral configuration, and this was supported by Neagi (1919) and Mills and Warren (1925). It was, however, Mills and Warren who gave the most conclusive evidence that the configuration is tetrahedral. Their evidence is based on the following argument. Compounds of the type *ab*C=C=C*ab* are resolvable since carbon is "tetrahedral" (see allenes, §6. V), and if nitrogen is also "tetrahedral", then the compound *ab*C=N=C*ab* should be resolvable, but will not be resolvable if the nitrogen is pyramidal. Mills and Warren prepared 4-carbethoxy-4′-phenylbispiperidinium-1:1′-spiran bromide, and resolved it. If the configuration of this molecule is VI, *i.e.*, a spiran, then it possesses no elements of symmetry, and hence will be resolvable; if the configuration is VII (*i.e.*, pyramidal), then it

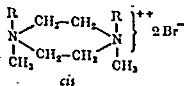


VI

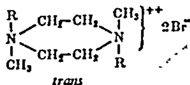


VII

will possess a vertical plane of symmetry, and hence will be optically inactive. Since the compound was resolved, the configuration must be tetrahedral, *i.e.*, VI. This tetrahedral configuration has been confirmed by physico-chemical studies (see §2b). More recently, Hanby and Rydon (1945) have shown that the diquaternary salts of dimethylpiperazine exhibit geometrical isomerism, and this is readily explained on the tetrahedral configuration of the four nitrogen valencies (*cf.* *cyclo*hexane, §11. IV).

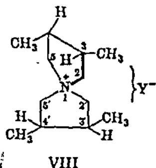


cis

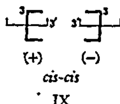


trans

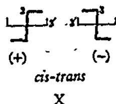
It has already been mentioned (§6. II) that McCasland and Proskow (1956) prepared a spiro-nitrogen compound which contained no plane or centre of symmetry, but was nevertheless optically inactive because it contained an alternating axis of symmetry. We shall now examine this compound (VIII; Y^- is the *p*-toluene-sulphonate ion) in more detail. This molecule can exist in four



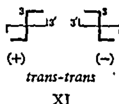
VIII



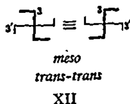
(+) (−)
cis-cis
IX



(+) (−)
cis-trans
X



(+) (−)
trans-trans
XI

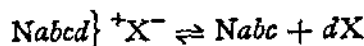


meso
trans-trans
XII

diastereoisomeric forms, three active and one *meso*. All four have been prepared, and are depicted as shown in IX, X, XI and XII. The co-axis of each spiran is assumed to be perpendicular to the

the heavy line, and a uniform orientation has been adopted: the front ring is always vertical, and the back horizontal ring with at least one substituent directed upwards and the *cis* ring placed at the back in the case of the *cis/trans* ring combination.

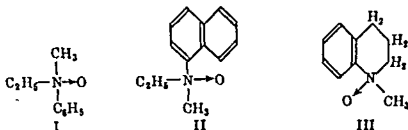
Racemisation of optically active quaternary ammonium salts is far more readily effected than that of carbon compounds containing an asymmetric carbon atom, *i.e.*, compounds of the type *Cabde*. The mechanism of the racemisation of the ammonium salts is believed to take place by dissociation into the amine, which then rapidly racemises (§2c):



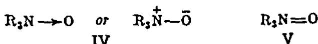
$$Nabcd\}^+X^- \rightleftharpoons Nabc + dX$$

Recombination of the racemised amine with dX results in the racemisation of the quaternary compound (see §4a).

§2b. Tertiary amine oxides. In tertiary amine oxides, *abc*NO, the nitrogen atom is joined to four different groups, and on the basis that the configuration is tetrahedral, such compounds should be resolvable. In 1908, Meisenheimer resolved ethylmethylphenylamine oxide, I, and this was then followed by the resolution of other amine oxides, *e.g.*, ethylmethyl-1-naphthylamine oxide, II, and kairoline oxide, III.



The evidence in favour of the structure IV as opposed to that of V is based on dipole moment measurements and on the fact that such compounds can be resolved. It should be noted that the



pyramidal structure would also account for the optical activity of these compounds as well as the tetrahedral. Consequently these compounds cannot be used as a criterion for the pyramidal or tetrahedral configuration of the nitrogen atom. However, by analogy with the quaternary ammonium salts, the configuration of amine oxides may be accepted as tetrahedral. Further evidence for this is as follows. The electronic configuration of nitrogen is $(1s^2)(2s^2)(2p^3)$. For nitrogen to be quinquevalent, the "valence state" will be derived from the arrangement $(1s^2)(2s)(2p^3)(3s)$. Now the amount of energy required to promote an electron from a $2s$ to a $3s$ orbital appears to be too large for it to occur, and consequently nitrogen is (apparently) never quinquevalent. The valence state of nitrogen is thus achieved by the *loss* of one $2s$ electron and then hybridisation of the $2s$ and $2p^3$ orbitals, *i.e.*, nitrogen becomes quadricovalent unielectrovalent, and the four bonds (sp^3 bonds) are arranged tetrahedrally. The charged nitrogen atom is isoelectronic with carbon, and so one can expect the formation of similar bonds. Furthermore, evidence obtained by an examination of the vibration frequencies of the ammonium ion indicates that the configuration of this ion is tetrahedral.

rapidly at right angles above and below the plane containing the groups a, b and c (see Fig. 1); II and III are the two extreme

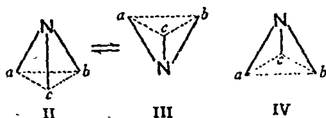


Fig. 6.1.

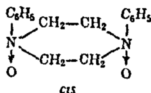
forms, and they are mirror images and not superimposable (IV is III "turned over", and it can be seen that IV is the mirror image of II). Thus this oscillation brings about very rapid optical inversion. This oscillation theory is supported by evidence obtained from the absorption spectrum of ammonia (Barker, 1929; Badger, 1930), and the frequency of the oscillation (and therefore the inversion) has been calculated to be $2 \cdot 3 \times 10^{10}$ per second (Cleeton *et al.*, 1934).

In the foregoing explanation for the racemisation of amines, it has been assumed that the nitrogen valency angles and the bond lengths change. This inversion of amines, however, is better represented as an "umbrella" switch of bonds, *i.e.*, the bond lengths remain unaltered and only the nitrogen valency angles change. This interpretation is more in keeping with the facts, *e.g.*, as the groups a, b and c increase in weight, the frequency of the inversion of the molecule decreases.

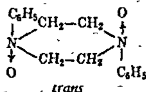
Theoretical calculations have shown that an optically active compound will not racemise spontaneously provided that the energy of activation for the change of one enantiomorph into the other is greater than 12–15 kg.cal./mole. The two forms, II and III, have been shown to be separated by an energy barrier of about 6 kg.cal./-mole, and consequently the two forms are readily interconvertible.

It has already been mentioned (§2b) that the electronic configuration of the nitrogen atom is $(1s^2)(2s^2)(2p^3)$. According to Hund's rule, electrons tend to avoid being in the same orbital as far as possible (see Vol. I, Ch. II). Thus, in ammonia and its derivatives, bonds are formed by pairing with the three single orbitals $2p_x$, $2p_y$ and $2p_z$. Since these are mutually at right angles, the configuration of the ammonia molecule will be a *trigonal pyramid*, *i.e.*, a pyramid with a triangular base, with the nitrogen atom situated at one corner. Oscillation of the nitrogen atom brings about inversion in the tertiary amines, Nabc. This picture of the configuration of the ammonia molecule, however, requires

Recently, Bennett and Glynn (1950) have obtained two geometrical isomers of 1:4-diphenylpiperazine dioxide; this is readily explained *on the tetrahedral configuration of nitrogen* (*cf.* §2a).

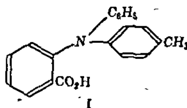


cis



trans

§2c. Amines. *If the tertiary amine molecule, Nabc, is planar,* it will be superimposable on its mirror image, and therefore cannot be optically active. All attempts to obtain tertiary amines in optically active forms have failed up to the present time, *e.g.*, Kipping and Salway (1904) treated secondary amines, *R·NH·R'*, with (±)-benzylmethylacetyl chloride; if the three valencies of the nitrogen atom are not planar, then the base will be a racemic modification, and on reaction with the acid chloride, the following four substituted amides should be formed: B_+A_+, B_-A_-, B_+A_-, B_-A_+, *i.e.*, a mixture of two pairs of enantiomorphs. Experiments carried out with, *e.g.*, methylaniline and benzylaniline gave *homogeneous* products. Meisenheimer *et al.* (1924) attempted to resolve *N*-phenyl-*N*-*p*-tolylanthranilic acid, I, and also failed. In view of



I

these failures, it would thus appear that the tertiary amine molecule is planar. Physico-chemical methods, *e.g.*, dipole moment measurements, infra-red absorption spectra studies, etc., have, however, shown conclusively that *the configuration of ammonia and of tertiary amines is tetrahedral.* Thus ammonia has been shown to have a dipole moment of 1·5D; had the molecule been planar, the dipole moment would have been zero. Furthermore, the *nitrogen valency angles in, e.g., trimethylamine* have been found to be 108°, thus again showing that the amine molecule is *not* planar. Why, then, cannot tertiary amines be resolved? Is it a question of experimental technique, or is there something inherent in the tertiary amine molecule that makes it impossible to be resolved? Meisenheimer (1924) explained the failure to resolve as follows. In the tertiary amine molecule, the nitrogen atom oscillates

rapidly at right angles above and below the plane containing the groups *a*, *b* and *c* (see Fig. 1); II and III are the two extreme

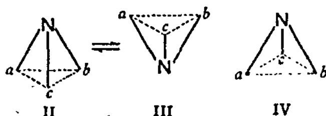


FIG. 6.1.

forms, and they are mirror images and not superimposable (IV is III "turned over", and it can be seen that IV is the mirror image of II). Thus this oscillation brings about very rapid optical inversion. This oscillation theory is supported by evidence obtained from the absorption spectrum of ammonia (Barker, 1929; Badger, 1930), and the frequency of the oscillation (and therefore the inversion) has been calculated to be $2{\cdot}3 \times 10^{10}$ per second (Cleeton *et al.*, 1934).

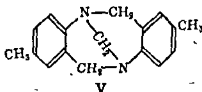
In the foregoing explanation for the racemisation of amines, it has been assumed that the nitrogen valency angles and the bond lengths change. This inversion of amines, however, is better represented as an "umbrella" switch of bonds, *i.e.*, the bond lengths remain unaltered and only the nitrogen valency angles change. This interpretation is more in keeping with the facts, *e.g.*, as the groups *a*, *b* and *c* increase in weight, the frequency of the inversion of the molecule decreases.

Theoretical calculations have shown that an optically active compound will not racemise spontaneously provided that the energy of activation for the change of one enantiomorph into the other is greater than 12–15 kg.cal./mole. The two forms, II and III, have been shown to be separated by an energy barrier of about 6 kg.cal./-mole, and consequently the two forms are readily interconvertible.

It has already been mentioned (§2b) that the electronic configuration of the nitrogen atom is $(1s^2)(2s^2)(2p^3)$. According to Hund's rule, electrons tend to avoid being in the same orbital as far as possible (see Vol. I, Ch. II). Thus, in ammonia and its derivatives, bonds are formed by pairing with the three single orbitals $2p_x$, $2p_y$ and $2p_z$. Since these are mutually at right angles, the configuration of the ammonia molecule will be a *trigonal pyramid*, *i.e.*, a pyramid with a triangular base, with the nitrogen atom situated at one corner. Oscillation of the nitrogen atom brings about inversion in the tertiary amines, N*abc*. This picture of the configuration of the ammonia molecule, however, requires

modification. The valency angles in ammonia have been shown to be approximately 107°. The deviation from the value of 90° (on the assumption that the bonds are pure $2p$ orbitals) is too great to be accounted for by repulsion between the hydrogen atoms. One explanation that has been offered is that the lone pair is not pure $(2s^2)$, but that " mixing " of them with the $2p$ electrons causes the bonds to open (see §13. I). This means that the nitrogen atom in tercovalent compounds has tetrahedral hybridisation, one orbital containing a lone pair of electrons (which have some p-character), and the other three orbitals one electron each. Thus the valency angles would be about 109·5° (*cf.* carbon, §4. II). As we have seen above, the valency angles in ammonia are 107°, and according to Mellish *et al.* (1954) this smaller value is due to the lone pair of electrons occupying a larger volume than the electrons of the N—H bonds, thus forcing the latter together to give an H—N—H angle less than the tetrahedral value.

In view of what has been said above, it appears that tertiary amines of the type N*abc* will never be resolved. Now, Kincaid and Henriques (1940), on the basis of calculations of the energy of activation required for the inversion of the amine molecule, arrived at the conclusion that tertiary amines are incapable of resolution because of the ease of racemisation, but if the nitrogen atom formed a part of a *ring* system, then the compound would be sufficiently optically stable to be isolated. This prediction was confirmed by Prelog and Wieland (1944), who resolved Troger's

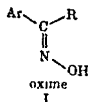


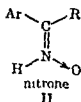
V

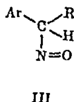
base, V, by chromatographic adsorption on D-lactose (*cf.* §10 vi. II). In this compound, the nitrogen is tervalent, but the frequency of oscillation has been brought to zero by having the three valencies of nitrogen as part of the ring system.

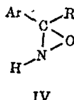
§2d. Oximes. In 1883, Goldschmidt found that benzil dioxime, C_6H_5·C(=NOH)·C(=NOH)·C_6H_5, could be converted into an isomeric form by boiling it in ethanolic solution; and then, in 1889, Meyer *et al.* isolated a third isomer of this compound. Beckmann, also in 1889, found that benzaldoxime existed in two isomeric forms, and from that time many aromatic oximes were shown to

exist in two isomeric forms. The existence of isomerism in aromatic oximes was first explained by structural isomerism, two of the following four structures corresponding to the two isomers (where R is an alkyl or an aryl group); II is the modern way of writing

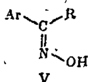


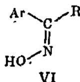






the nitrone structure (originally, it was written with quinquevalent nitrogen, the nitrogen being linked to the oxygen by a double bond). Hantzsch and Werner (1890), however, suggested that the isomerism of the oximes was geometrical and not structural. According to these authors, nitrogen is tervalent (in oximes), and is situated at one corner of a tetrahedron with its three valencies directed towards the other three corners; consequently the three valencies are not coplanar (*cf.* tertiary amines). These authors also assumed that there is no free rotation about the C=N double bond (*cf.* §2. IV), and therefore proposed configurations V and VI for the two isomers:





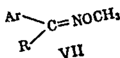
Many facts are in favour of geometrical isomerism, *e.g.*,

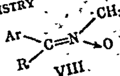
(i) If Ar=R, then isomerism disappears.

(ii) III and IV would be optically active; this is not found to be so in practice.

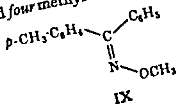
(iii) Absorption spectra measurements show that the two isomers have identical structures.

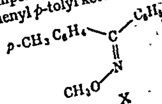
As pointed out above, Hantzsch and Werner chose structure I as the formula for the oximes, but examination of II shows that this would also satisfy the requirements for geometrical isomerism; structure I was chosen because oximes were known to contain the group >C=NOH. Later work, however, has shown that the problem is not so simple as this; methylation of an oxime (with methyl sulphate) usually produces a mixture of two compounds, one of which is the *O*-methyl ether, VII, and the other the *N*-methyl

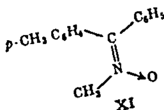


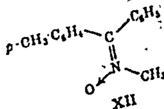


ether, VIII. These two are readily distinguished by the fact that on heating with hydriodic acid, VII gives methyl iodide, whereas VIII gives methylamine. Thus, Semper and Lichtenstadt (1918) obtained *four* methyl derivatives of phenyl *p*-tolyl ketoxime, IX–XII.

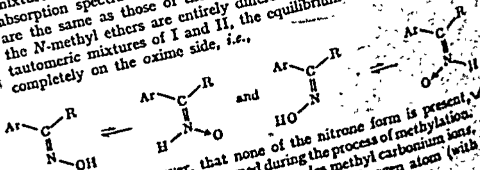






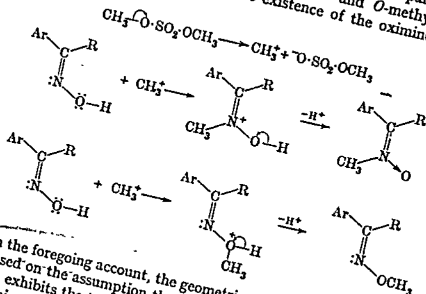


On treatment with concentrated hydriodic acid, two of these compounds gave methyl iodide, and therefore correspond to the *O*-methyl derivatives, IX and X; the other two compounds gave methylamine, and therefore correspond to the *N*-methyl derivatives, XI and XII. Thus it appears that oximes can exist in forms I and II. Brady (1916) considered that oximes in solution are a tautomeric mixture of I and II (*oximino-nitrone diad system*). Ultraviolet absorption spectra studies show that the spectra of the oximes are the same as those of the *O*-methyl ethers, whereas those of the *N*-methyl ethers are entirely different. Hence, if oximes are tautomeric mixtures of I and II, the equilibrium must lie almost completely on the oxime side, *i.e.*,

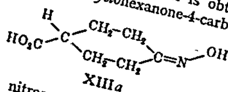


It is possible, however, that none of the nitrone form is present, but its methyl derivative is formed during the process of methylation. If we assume that methyl sulphate provides methyl carbonium ions, then it is possible that these ions attack the nitrogen atom (with

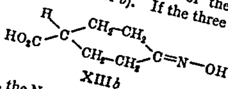
its lone pair) or the oxygen atom (with its two lone pairs). This would result in the formation of the *N*- and *O*-methyl ethers, without having to postulate the existence of the oximino-nitrone tautomeric system.



In the foregoing account, the geometrical isomerism of the oximes is based on the assumption that the nitrogen atom, in the oximino-form, exhibits the trigonal pyramidal configuration. Further proof for this configuration is obtained from the examination of the oxime of *cyclo*hexanone-4-carboxylic acid (XIII*a* or *b*). If the three



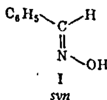
XIII*a*

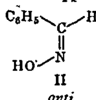


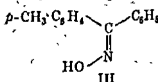
XIII*b*

nitrogen valencies are non-planar (*i.e.*, the N—O bond is not collinear with the C=N double bond), the configuration is XIII*a*, and it will therefore be optically active. If, however, the three nitrogen valencies are coplanar and symmetrically placed, then the configuration will be XIII*b*, and this will not be optically active, since it possesses a plane of symmetry. Mills and Bain (1910) prepared this oxime and resolved it; hence its configuration must be XIII*a*. This is readily explained on the modern theory of valency. The three valencies of 3-covalent nitrogen are the three $2p$ orbitals (§2c). If the $2p_x$ orbital of the nitrogen atom forms a σ bond with one of the sp^2 bonds of the carbon atom, and if the $2p_z$ orbital forms a π bond with that of the p_z orbital of the carbon atom, then the OH group is bonded with the $2p_y$ orbital of the nitrogen atom and consequently lies along a line (approximately) perpendicular to the direction of the "double" bond.

§2e. Nomenclature of the oximes. In oxime chemistry the terms *syn* and *anti* are used instead of the terms *cis* and *trans*. When dealing with aldoximes, the *syn*-form is the one in which both the hydrogen atom and the hydroxyl group are on the same side; when these groups are on opposite sides, the configuration is

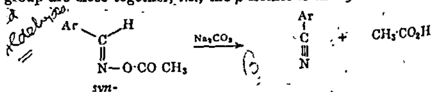


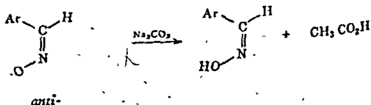




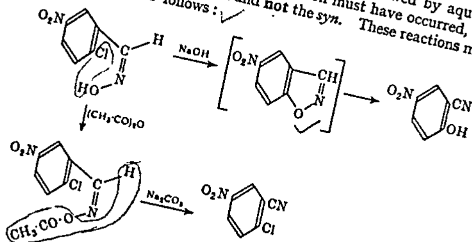
anti. Thus I is *syn*- and II is *anti*-benzaldoxime. With ketoximes, the prefix indicates the spatial relationship between the *first* group named and the hydroxyl group (*cf.* §4. IV). Thus III may be named as *syn*-*p*-tolyl phenyl ketoxime or *anti*-phenyl *p*-tolyl ketoxime.

§2f. Determination of the configuration of aldoximes. As we have seen, aromatic aldoximes can be obtained in two geometrical isomeric forms, the *syn* and the *anti*. Aliphatic aldoximes, however, appear to occur in one form only, and this is, apparently, the *anti*-form. The problem, then, with aromatic aldoximes, is to assign configurations to the stereoisomeric forms. The two forms (of a given aldoxime) resemble each other in many ways, but differ very much in the behaviour of their acetyl derivatives towards aqueous sodium carbonate. The acetyl derivative of one isomer regenerates the aldoxime; this form is known as the α-isomer. The other isomer, however, eliminates a molecule of acetic acid to form an aryl cyanide; this form is known as the β-isomer. Hantzsch and Werner (1890) suggested that the β-form readily eliminates acetic acid because the hydrogen atom and the acetoxy-group are close together, *i.e.*, the β-isomer is the *syn*-form; thus:

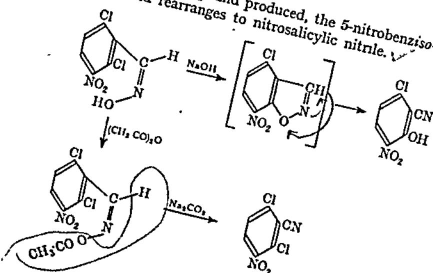




Such a view, however, is contrary to many experimental results (*cf.* §5 xi. IV); *e.g.*, Brady and Bishop (1925) found that only one of the two isomers of 2-chloro-5-nitrobenzaldoxime readily gave ring closure on treatment with sodium hydroxide. It therefore follows that this form is the *anti*-isomer (*cf.* method of cyclisation, §5 i. IV). It was also found that it was this isomer that gave the cyanide on treatment with acetic anhydride followed by aqueous sodium carbonate. Thus *anti*-elimination must have occurred, *i.e.*, the β-isomer is the *anti*-form and **not** the *syn*. These reactions may be formulated as follows:

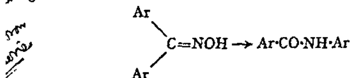


Actually, the ring compound produced, the 5-nitrobenz*iso*-oxazole, is unstable, and rearranges to nitrosalicylic nitrile.



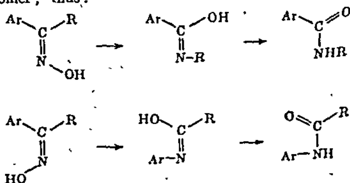
In a similar manner, Meisenheimer (1932) found that of the two isomeric 2:6-dichloro-3-nitrobenzaldoximes, it was the *anti*-isomer that gave ring closure, and was also the one that gave the cyanide. Hence, if *anti*-elimination is used as the criterion for these reactions, the configurations of the *syn*- and *anti*-forms can be determined. It might be noted here, in passing, that since the *syn* form was originally believed to form the cyanide, the configurations of the isomers in the literature up to 1925 (*i.e.*, before Brady's work) are the reverse of those accepted now.

§2g. Determination of the configuration of ketoximes. The configurations of ketoximes have been mainly determined by means of the **Beckmann rearrangement** (1886). Aromatic ketoximes, *i.e.*, ketoximes containing at least one aromatic group, occur in two forms; aliphatic ketoximes appear to occur in one form only. When treated with certain acidic reagents such as sulphuric acid, acid chlorides, acid anhydrides, phosphorus pentachloride, etc., ketoximes undergo a molecular rearrangement, resulting in the formation of an acid amide:

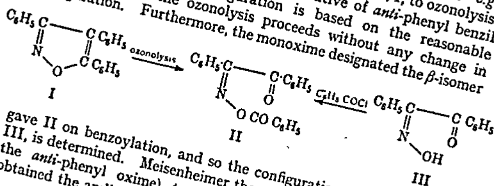


This rearrangement is known as the *Beckmann rearrangement* or *Beckmann transformation.* The best method is to treat an ethereal solution of the oxime with phosphorus pentachloride at a temperature below − 20°. On the other hand, Hornung *et al.* (1952) have found that a very good method for effecting the Beckmann rearrangement is to heat the oxime in polyphosphoric acid at 95° to 130°.

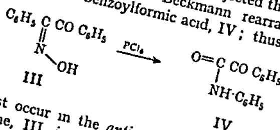
Hantzsch (1891) suggested that the course of the rearrangement indicated the configuration of the oxime, and assumed that the *syn*-exchange of groups occurred since they were closer together in this isomer; thus:



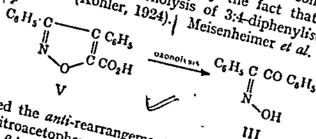
This, again, was shown experimentally to be the reverse, *i.e.*, it is the *anti*-rearrangement that occurs, and not the *syn*; *e.g.*, Meisenheimer (1921) subjected triphenyl*iso*-oxazole, I, to ozonolysis, and thereby obtained the benzoyl-derivative of *anti*-phenyl benzil monoxime, II. This configuration is based on the reasonable assumption that the ozonolysis proceeds without any change in configuration. Furthermore, the monoxime designated the β-isomer



gave II on benzoylation, and so the configuration of the β-isomer, III, is determined. Meisenheimer then subjected this β-oxime (*i.e.*, the *anti*-phenyl oxime) to the Beckmann rearrangement, and obtained the anilide of benzoylformic acid, IV; thus the exchang



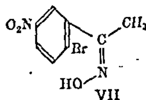
of groups must occur in the *anti*-position. The configuration of the β-monoxime, III, is confirmed by the fact that it may be obtained directly by the ozonolysis of 3:4-diphenyl*iso*-oxazole-5-carboxylic acid, V (Kohler, 1924). Meisenheimer *et al.* (1925) also

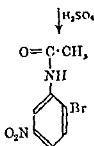


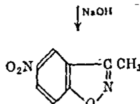
demonstrated the *anti*-rearrangement as follows. The α-oxime of 2-bromo-5-nitroacetophenone is unaffected by sodium hydroxide, whereas the β-isomer undergoes ring closure to form 3-methyl-5-nitrobenz*iso*-oxazole; thus the α-oxime is the *syn*-methyl isomer VI, and the β-oxime the *anti*-methyl isomer VII. When treated with sulphuric acid or phosphorus pentachloride, the α-oxime underwent the Beckmann rearrangement to give the *N*-substituted acetamide; thus the exchange occurs in the *anti*-positions.

Further evidence for the *anti*-exchange of groups in the Beckmann rearrangement has been obtained by studying the behaviour

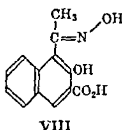


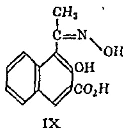






of compounds exhibiting restricted rotation about a single bond, *e.g*, Meisenheimer *et al.* (1932) prepared the two-isomeric oximes of 1-acetyl-2-hydroxynaphthalene-3-carboxylic acid, VIII and IX,



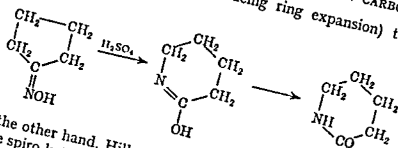


and of these two forms only one was resolvable. This resolvable isomer must therefore be IX, since asymmetry due to restricted rotation is possible only with this form (*cf.* §3. V). Meisenheimer found that the ethyl ester of IX, on undergoing the Beckmann rearrangement, gave the amide Ar·CO·NH·CH_3 (where Ar is the naphthalene part of the molecule), whereas the ethyl ester of VIII gave the amide CH_3·CO·NH·Ar. These results are in agreement with the *anti*-exchange of groups in each case.

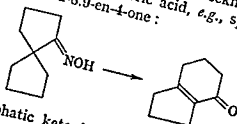
Thus the evidence is all in favour of the *anti*-exchange of groups in the Beckmann rearrangement, and hence by using this principle, the Beckmann rearrangement may be used to determine the configuration of ketoximes.

An interesting application of the Beckmann rearrangement is in the formation of heterocyclic rings, *e.g.*, when *cyclo*pentanone-oxime is subjected to the Beckmann rearrangement, the nitrogen

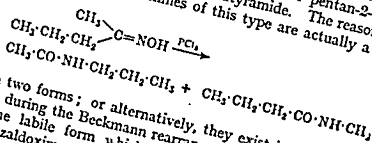
atom enters the ring (thus producing ring expansion) to form 2-piperidone.



On the other hand, Hill *et al.* (1956) have shown that the oximes of some spiro-ketones undergo abnormal Beckmann rearrangements in the presence of polyphosphoric acid, *e.g.*, spiro-[4:4]-nonanone-1-oxime gives hydrind-8.9-en-4-one:

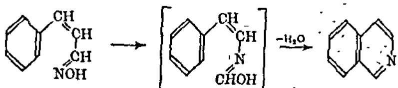


Although aliphatic ketoximes are not known in two isomeric forms, some may produce two products when subjected to the Beckmann rearrangement, *e.g.*, the oxime of pentan-2-one gives *N*-propylacetamide and *N*-methylbutyramide. The reason for this is uncertain; possibly oximes of this type are actually a mixture

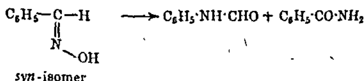


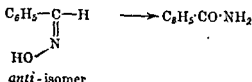
of the two forms; or alternatively, they exist in one stable form which, during the Beckmann rearrangement, is partially converted into the labile form which then undergoes the rearrangement (*cf.* benzaldoxime, below).

Whereas the majority of ketoximes undergo the Beckmann rearrangement, it appears that few aldoximes do so. In an attempt to prepare quinoline by the dehydration of cinnamaldoxime with phosphorus pentoxide, Bamberger and Goldschmidt (1894) actually obtained *iso*quinoline; the formation of the latter compound and not the former can only be reasonably explained on the assumption that the oxime first undergoes the Beckmann rearrangement, and the rearranged product then undergoes ring closure to form *iso*quinoline. Recently, Hornung *et al.* (1952) have shown th



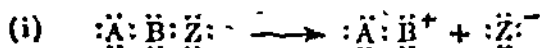
aldoximes can be made to undergo the Beckmann rearrangement under the influence of polyphosphoric acid, *e.g.*, *syn*-benzaldoxime gives a mixture of formanilide and benzamide, the latter being



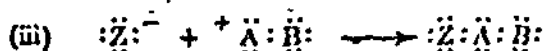


produced by the conversion of the *syn*-form into the *anti*; *anti*-benzaldoxime gives benzamide only. These results are in agreement with the configurations obtained by other methods (see §2*l*).

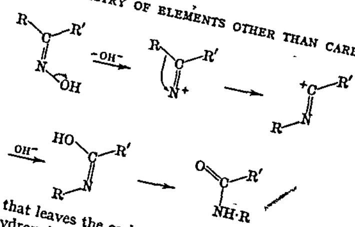
§2h. Mechanism of the Beckmann rearrangement. The mechanism of the Beckmann rearrangement is *still not* certain; one mechanism favoured by many is that based on the Whitmore mechanism, which may be formulated in the following general terms (see also Vol. I, Ch. VII):





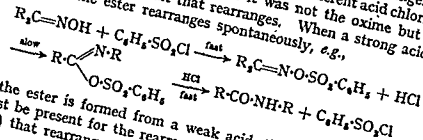


If B has a greater electron-affinity than A, the electrons rearrange as in (ii), *the shift in the electron pair including the atom or group which it holds.* Then recombination with Z occurs to form the rearranged product ZAB (reaction iii). Thus, on this basis, the Beckmann rearrangement may be formulated as follows:



The radical that leaves the carbon atom is the one on the *opposite* side to the hydroxyl group, *i.e.*, there is an *anti*-exchange of groups, and this leads to the "inversion" of the configuration of the nitrogen atom. It is interesting to note in connection with the above mechanism that it is in keeping with Beckmann's idea that the rearrangement occurred by a direct interchange of groups, the reagent acting as a catalyst.

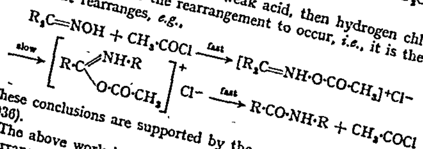
The mechanism of the rearrangement, however, is not so simple as this. Kuhura (1926) investigated the speed of the rearrangement of benzophenone oxime under the influence of different acid chlorides, and arrived at the conclusion that it was not the oxime but the acyl derivative or its salt that rearranges. When a strong acid is used, then the ester rearranges spontaneously, *e.g.*,



$$R_2C{=}NOH + C_6H_5{\cdot}SO_2Cl \xrightarrow{\text{fast}} R_2C{=}N{\cdot}O{\cdot}SO_2{\cdot}C_6H_5 + HCl$$

$$\xrightarrow{\text{slow}} R{\cdot}C\begin{matrix}{=}N{\cdot}R\\ \diagdown O{\cdot}SO_2{\cdot}C_6H_5\end{matrix} \xrightarrow[\text{fast}]{HCl} R{\cdot}CO{\cdot}NH{\cdot}R + C_6H_5{\cdot}SO_2Cl$$

If the ester is formed from a weak acid, then hydrogen chloride must be present for the rearrangement to occur, *i.e.*, it is the salt (ion) that rearranges, *e.g.*,



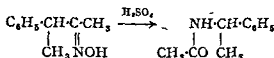
$$R_2C{=}NOH + CH_3{\cdot}COCl \xrightarrow{\text{fast}} [R_2C{=}NH{\cdot}O{\cdot}CO{\cdot}CH_3]^+Cl^-$$

$$\xrightarrow{\text{slow}} \left[R{\cdot}C\begin{matrix}{=}NH{\cdot}R\\ \diagdown O{\cdot}CO{\cdot}CH_3\end{matrix}\right]^+ Cl^- \xrightarrow{\text{fast}} R{\cdot}CO{\cdot}NH{\cdot}R + CH_3{\cdot}COCl$$

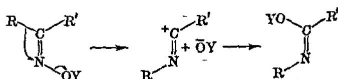
These conclusions are supported by the work of Chapman (1933–1936).

The above work indicates that it is not the oxime itself which rearranges, but some intermediate; it does not explain, however, how this intermediate rearranges. Now it has been shown that when the migrating alkyl group contains an asymmetric carbon

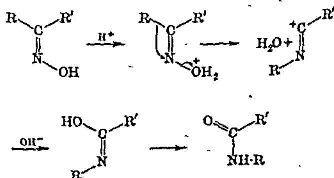
atom, there is no loss of optical activity during the rearrangement, *e.g.*, Kenyon *et al.* (1946) have shown that when (+)-α-phenylethyl methyl ketoxime is treated with sulphuric acid, the product, *N*-α-phenylethylacetamide, is obtained in almost 100 per cent.



optical purity. This indicates that the rearrangement occurs intramolecularly, since if the alkyl group separated as a carbanion the optical activity would presumably be lost. Furthermore, since the oxime itself does not rearrange, but the intermediate does, it seems reasonable to conclude that the hydroxyl group, as a hydroxyl ion, cannot separate from the nitrogen atom, whereas the group OY^- (where Y is some acid radical) can separate readily. Experiments have shown that benzophenone oxime picryl ether rearranges more rapidly in polar solvents than in non-polar (Chapman, 1934). This is strong evidence that the rate-determining step in the rearrangement is the ionisation of the intermediate. Thus the mechanism may be postulated as follows:



When the reagent used for effecting the rearrangement is a strong inorganic acid, the mechanism may then be:



The carbonium ion may combine with the hydroxyl ion which was originally eliminated (as water), or with a hydroxyl ion provided by the solution. Evidence for the latter is afforded by the experiments of Brodski *et al.* (1941), who found that when benzophenone oxime was subjected to the Beckmann rearrangement in water

containing the isotope ^{18}O, the benzanilide obtained contained some of this isotope.

Pearson *et al.* (1955) have shown that the Beckmann rearrangement of acetophenone oxime is sterically accelerated (§6a. III) by the presence of *o*-substituents.

Stephen *et al.* (1956) have shown that *one* molecule of phosphorus pentachloride, phosphoryl chloride, thionyl chloride, or benzenesulphonyl chloride rearranges *two* molecules of the ketoxime to yield the corresponding amide and imidoyl chloride in approximately equimolecular amounts, *e.g.*,

$$2R_2C{=}NOH + PCl_5 \longrightarrow R{\cdot}CO{\cdot}NH{\cdot}R + R{\cdot}CCl{=}N{\cdot}R + POCl_3 + HCl$$

It has also been shown that hydrogen chloride is essential during the rearrangement, but that it does not itself cause the rearrangement of the oxime. On the basis of these results, Stephen *et al.* have proposed the following mechanism for the Beckmann rearrangement of ketoximes. The reagent first produces some acid amide and imidoyl chloride, and the latter then dehydrates unchanged ketoxime to the anhydride which then reacts as shown:

$$2R_2C{=}NOH \xrightarrow{-H_2O} \underset{\text{anhydride}}{(R_2C{=}N{-})_2O} \xrightarrow{HCl} \underset{\text{anhydride salt}}{[R_2C{=}\overset{+}{N}H{\cdot}O{\cdot}N{=}CR_2]Cl^-}$$

$$\longrightarrow \underset{\text{ketoxime imidate}}{R{\cdot}N{=}CR{\cdot}O{\cdot}N{=}CR_2} \longrightarrow \underset{\text{imidoyl anhydride}}{(R{\cdot}N{=}CR{-})_2O}$$

$$\xrightarrow{HCl} R{\cdot}CCl{=}N{\cdot}R + R{\cdot}CO{\cdot}NH{\cdot}R$$

It is also suggested that other reagents which effect the Beckmann rearrangement may function as dehydrating agents for the formation of the ketoxime anhydride.

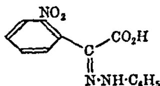
When a *trace* of the reagent is used, a large yield of amide is obtained. The mechanism is believed to be the same as that given above, provided that in the initial stage there is sufficient to form a trace of the ketoxime anhydride in the presence of hydrogen chloride. Rearrangement of the anhydride will now take place as above with the formation of the imidoyl chloride which can then dehydrate ketoxime to anhydride, itself being converted into the amide:

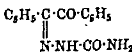
$$2R_2C{=}NOH + R{\cdot}CCl{=}N{\cdot}R \longrightarrow (R_2C{=}N{-})_2O + R{\cdot}CO{\cdot}NH{\cdot}R + HCl$$

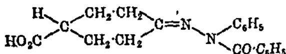
Thus the yield of amide increases at the expense of the imidoyl chloride.

§2l. Stereoisomerism of some other tervalent nitrogen compounds containing a double bond. There are several other

types of compounds besides the oximes in which the nitrogen atom is linked by a double bond. The other atom joined by this double bond may be a carbon atom (as in the oximes), or another nitrogen

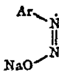


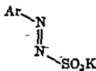


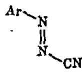


atom, and in both cases stereoisomerism is possible; *e.g.*, Krause (1890) obtained two isomeric forms of the phenylhydrazone of *o*-nitrophenylglyoxylic acid, I, and Hopper (1925) isolated two isomers of the monosemicarbazone of benzil, II. *Mills* and *Bain* (1914) resolved III; this is resolvable because of the non-planar configuration of the three nitrogen valencies (*cf.* the oximes, §2d).

Many cases of geometrical isomerism are known in which the two forms are due to the presence of a nitrogen-nitrogen double bond. Examples of this type which have been most extensively studied are the diazoates, IV, the diazosulphonates, V, and the diazocyanides, VI (see Vol. I, Ch. XXIV, for an account of these compounds).

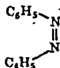




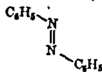


IV *syn*-form V *anti*-form VI *anti*-form

Azobenzene is also an example of this type, and according to Hartley (1938), "ordinary" azobenzene is the *anti*-form.

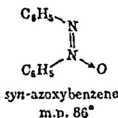


syn-azobenzene m.p. 71·4°

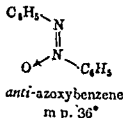


anti-azobenzene m.p. 68°

Azoxybenzene (in which one nitrogen atom is tercovalent and the other quadricovalent) also exists in two geometrical isomeric forms, the *anti*-isomer being "ordinary" azoxybenzene.

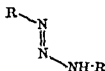


syn-azoxybenzene
m.p. 86°



anti-azoxybenzene
m.p. 36°

Recently, Le Fèvre *et al.* (1951) have measured the dipole moments and the ultraviolet absorption spectra of a number of triazens, and have concluded that these compounds exist in the *anti*-configuration about the nitrogen-nitrogen double bond, *i.e.*, the configuration is:



These authors also believe that this *anti*-form is converted into an equilibrium mixture of the *anti*- and *syn*-forms when exposed to sunlight.

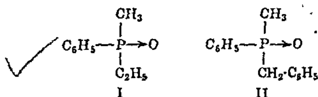
§3. STEREOCHEMISTRY OF PHOSPHORUS COMPOUNDS

Nitrogen, as we have seen, can exhibit covalencies of 3 and 4; phosphorus (and arsenic), however, can exhibit covalencies of 3, 4, 5 and 6, and consequently gives rise to more possible configurations than nitrogen.

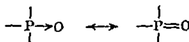
§3a. Tercovalent phosphorus compounds. Since the electronic configuration of phosphorus is $(1s^2)(2s^2)(2p^6)(3s^2)(3p^3)$, it might be expected that suitable tercovalent compounds, R_3P, could be resolved, since the phosphorus atom would occupy one corner of a tetrahedron, *i.e.*, the configuration would be a trigonal pyramid (*cf.* §2c). This configuration is to be expected on the basis that in the tercovalent state, phosphorus uses only the three $3p$ electrons; this is supported by the fact that the valency angle in phosphine is approximately 94°, and in trimethylphosphine the angle is $101 \pm 3°$. No tertiary phosphines, however, have yet been resolved, and the reason for this appears to be the same as for tertiary amines,

viz., that the phosphorus atom is in a state of oscillation. Calculation has shown that the frequency of this oscillation in phosphine is 5×10^6; this is slower than that of *nitrogen* ($2{\cdot}3 \times 10^{10}$), and if it could be brought to zero, then tertiary phosphines would be resolvable. Increasing the weight of the groups slows down the oscillation in phosphorus compounds, *e.g.*, replacement of the three hydrogen atoms by deuterium atoms changes the frequency to 6×10^3. It seems possible, therefore, that very large groups might produce phosphines which would be resolvable; and if not zero in these compounds, the oscillation certainly can be expected to be zero in ring compounds (*cf.* nitrogen, §2c). Thus, if chemical difficulties can be overcome, tercovalent phosphorus compounds would be resolvable (see also §4c).

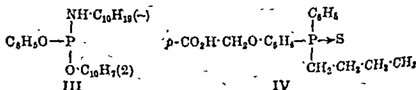
§3b. Quadricovalent (or quinquevalent) phosphorus compounds. The earliest phosphorus compounds to be resolved were the phosphine oxides, *e.g.*, Meisenheimer *et al.* (1911) resolved ethylmethylphenylphosphine oxide, I, and benzylmethylphenylphosphine oxide, II. Recent measurements of the P—O (and



As—O) bond length indicate that this bond has some double bond character; hence this bond is possibly best represented as a resonance hybrid.

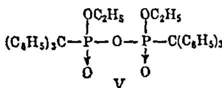


Kipping (1911) obtained two optically active forms of the *N*-(—)-menthyl derivative of 2-naphthylphenylphosphoramidic acid, III, and Davies and Mann (1944) resolved *n*-butylphenyl-*p*-carboxymethoxyphenylphosphine sulphide, IV.



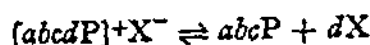
Another interesting phosphorus compound from the point of view of optical isomerism is ethyl triphenylmethylpyrophosphonate, V.

If the two phosphorus atoms are asymmetric, then V contains two similar asymmetric carbon atoms, and so its structure corresponds to the molecule *Cabd*·*Cabd*. Thus there will be one racemic modi-



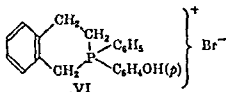
fication (composed of the pair of enantiomorphs) and one *meso*-form (*cf.* §7d. II). Hatt (1933) obtained two forms of compound V; both were inactive and so correspond to the racemic modification and the *meso*-form, but it was not possible to tell which was which.

Many attempts have been made to resolve quaternary phosphonium compounds, but until recently, all these attempts failed. This failure is attributed to the occurrence in solution of a "dissociation-equilibrium", which causes very rapid racemisation (see §4a).

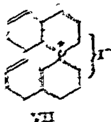


$$[abcdP]^+X^- \rightleftharpoons abcP + dX$$

The earlier attempts to resolve phosphonium compounds were always carried out on compounds containing at least one alkyl group; consequently dissociation in solution could occur, thereby resulting in racemisation. Holliman and Mann (1947) overcame this difficulty by preparing a much more stable type of phosphonium compound; these workers prepared a salt in which the phosphorus atom was in a ring, *viz.*, 2-phenyl-2-*p*-hydroxyphenyl-1:2:3:4-tetrahydro-*iso*phosphinolinium bromide, VI, and resolved it. The



resolution of 4-covalent compounds of phosphorus does not prove that the phosphorus atom has a tetrahedral configuration; it only proves that the phosphorus atom cannot be in the same plane as the other four groups attached to it. Mann *et al.* (1955), however, have now synthesised P-spiro-bis-1:2:3:4-tetrahydro-*iso*phosphinolinium iodide (VII) and resolved it into (+)- and (−)-forms which have high optical stability. The phosphorus atom is not asymmetric in this compound; it is the tetrahedral

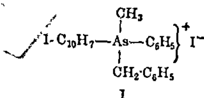


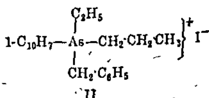
disposition of the four valencies which produces the dissymmetric cation (*cf* nitrogen, §2a; see also §4b).

§4. STEREOCHEMISTRY OF ARSENIC COMPOUNDS

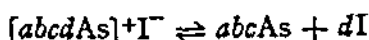
Arsenic, like phosphorus, can exhibit covalencies of 3, 4, 5 and 6; consequently these two elements show a great similarity to each other, and differ from nitrogen which has a maximum covalency of 4.

§4a. Quadricovalent arsenic compounds. The first resolution of an arsonium compound was carried out by Burrows and Turner (1921). These workers obtained a solution of benzylmethyl-



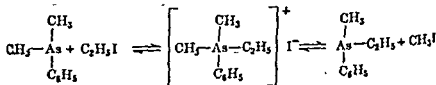


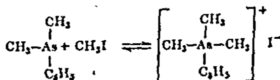
1-naphthylphenylarsonium iodide, I, that had a rotation of $+12°$, but racemised rapidly (in solution). Similarly, Kamai (1933) isolated the (+)-form of benzylethyl-1-naphthyl-*n*-propylarsonium iodide, II, which also racemised rapidly in solution. This rapid racemisation is believed to be due to a "dissociation-equilibrium" in solution. This explanation was suggested by Pope and Harvey (1901) to account for the racemisation of certain ammonium salts, but definite evidence for this theory was provided by Burrows and Turner (1921) in their work on arsonium salts. If this dissociation-equilibrium occurs, then in solution there will be:



$$[abcd\text{As}]^+\text{I}^- \rightleftharpoons abc\text{As} + d\text{I}$$

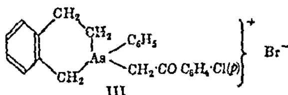
Burrows and Turner showed that when dimethylphenylarsine is treated with ethyl iodide, the expected ethyldimethylphenylarsonium





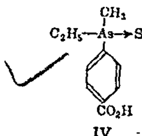
iodide is obtained, but at the same time a considerable amount of trimethylphenylarsonium iodide is also formed. These results are readily explained by the dissociation-equilibrium theory.

Since all the arsonium compounds investigated contained at least one alkyl group, Holliman and Mann (1943) prepared an arsonium compound with the arsenic atom in a ring, in the hope of stabilising the compound (*cf.* phosphorus, §3b). These authors prepared 2-*p*-chlorophenacyl-2-phenyl-1:2.3:4-tetrahydro-*iso*arsinolinium bromide, III, resolved it, and found that it did not racemise in solution at room temperature.

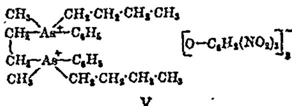


III

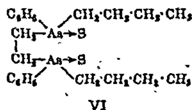
Although phosphine oxides of the type *abc*PO have been resolved (§3b), similar arsine oxides have not; the reason for this is obscure. On the other hand, arsine sulphides have been resolved, *e.g.*, Mills and Raper (1925) resolved *p*-carboxyphenylmethylethylarsine sulphide, IV.



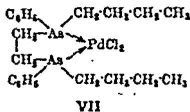
IV



V



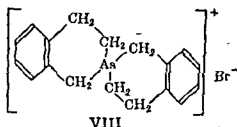
VI



VII

Chatt and Mann (1939) prepared ethylene-1:2-bis(*n*-butylmethylphenylarsonium) picrate, V, ethylene-1:2-bis(*n*-butylphenylarsine sulphide), VI, and ethylene-1:2-bis(*n*-butylphenylarsine)-dichloropalladium, VII, and obtained each compound in two forms. Each of these compounds is of the type *Cabd*·*Cabd*, and hence each should exist in one racemic modification and one *meso*-form (*cf.* §7d. II). As has already been stated, two forms of each were isolated; both were inactive, but the authors had no evidence for deciding which was which.

It has already been pointed out above that Holliman and Mann



VIII

prepared the optically stable arsonium compound III. These authors, in 1945, also resolved an arsonium compound of the spiran type, *viz*, As-spiro-bis-1:2:3:4-tetrahydro-*iso*arsinolinium bromide, VIII. This does not contain an asymmetric arsenic atom; the optical activity is due to the asymmetry of the molecule (the two rings are perpendicular to each other), and this is evidence that the four valencies of arsenic are arranged tetrahedrally (see also §4b).

§4b. Tercovalent arsenic compounds. The electronic configuration of arsenic is $(1s^2)(2s^2)(2p^6)(3s^2)(3p^6)(3d^{10})(4s^2)(4p^3)$. Thus the configuration of tercovalent arsenic compounds will be a trigonal pyramid (*cf.* phosphorus, §3a). Physico-chemical evidence (X-ray analysis, spectroscopy and electron diffraction) has shown that in tercovalent compounds the arsenic atom is at the apex of a tetrahedron, and that the intervalency angle is $100 \pm 4°$. It has also been shown that the arsenic is in a state of oscillation, the frequency of this oscillation through the plane of the three hydrogen atoms in arsine being 16×10^4. This is slower than that of phosphorus (5×10^6), and very much slower than that of nitrogen ($2{\cdot}3 \times 10^{10}$). Thus, preventing the oscillation of the arsenic atom, possibly by attachment to very large groups, should lead to the isolation of optically active tercovalent compounds. So far, however, all attempts to resolve compounds of the type *Asabc* have failed (*cf.* nitrogen and phosphorus). On the other hand, tercovalent arsenic compounds in which arsenic has two of its valencies occupied in a ring compound have been resolved; the ring structure prevents

oscillation of the arsenic atom (*cf.* Tröger's base, §2c). Thus Lesslie and Turner (1934) resolved 10-methylphenoxarsine-2-carboxylic acid, I. These authors suggested that the asymmetry of the molecule is due to the presence of a folded structure about the O—As axis, as well as the asymmetry due to the presence of an asymmetric arsenic atom (see structure II). This molecule and its mirror

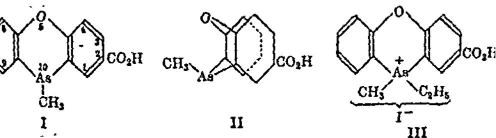
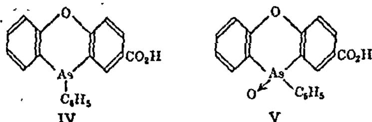


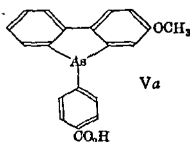
image are not superimposable. It might be noted, however, that the position of the methyl group with respect to the O—As axis is uncertain (*cf.* the arsanthrens, below). This folded structure is reasonable in view of the fact that the valency angle of oxygen is also approximately 100°; if the molecule were planar, then the valency angles of both arsenic and oxygen would be in the region of 120°, which is a very large increase from the normal valency angle. When each enantiomorph of II is treated with ethyl iodide, the *same* racemised product is obtained. This is due to the fact that when the arsonium compound, III, is formed, the asymmetric quaternary arsenic atom is racemised owing to the dissociation-equilibrium.



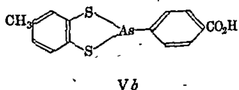
Lesslie and Turner (1936) also resolved 10-phenylphenoxarsine-2-carboxylic acid, IV. This compound was very stable, and oxidation to the arsine oxide, V, gave a completely racemised product.

Campbell *et al.* (1956) have resolved some substituted 9-arsafluorenes, *e.g.*, 9-*p*-carboxyphenyl-2-methoxy-9-arsafluorene (V*a*). Campbell (1956) has also resolved 2-*p*-carboxyphenyl-5-methyl-1:3-dithia-2-arsaindane (V*b*). This compound is optically stable in

chloroform solution, but is racemised in aqueous sodium hydroxide. Campbell believes that this racemisation is due to the fission of the As—S bonds by aqueous alkali, and subsequent reversal of the

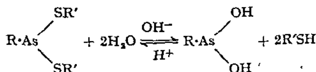


Va



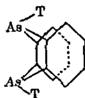
Vb

reaction by acid, a type of behaviour observed in triaryl thioarsenites (Klement *et al.*, 1938). Furthermore, Cohen *et al.* (1931) have shown that in sodium hydroxide solution, alkyl thioarsenites exist in equilibrium with thiol and arsenoxide:

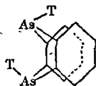


$$R{\cdot}As(SR')_2 + 2H_2O \underset{H^+}{\overset{OH^-}{\rightleftharpoons}} R{\cdot}As(OH)_2 + 2R'SH$$

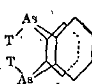
Chatt and Mann (1940) prepared 5:10-di-*p*-tolyl-5:10-dihydroarsanthren, and pointed out that if the valency angle of arsenic remains constant at its normal angle (of approximately 100°), then the structure will be folded, and consequently the three geometrical isomers, VI, VII and VIII, are apparently possible (*T* represents the *p*-tolyl group). Chatt and Mann also pointed out that evidence



VI



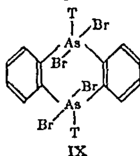
VII



VIII

obtained from models constructed to scale showed that the two *p*-tolyl radicals (T) in VIII would almost be coincident, and hence this isomer cannot exist. These authors isolated two optically inactive forms, but were unable to say which was which. When each compound was treated with bromine, both gave the *same* tetrabromide which, on hydrolysis, gave only one tetrahydroxide. The loss of isomerism in the tetrabromide (and in the tetrahydroxide) may be explained as follows. Bromination of VI and VII converts tercovalent arsenic into quinquecovalent arsenic, and in the latter

state the ring valency angles of the arsenic become 120°, and so the arsanthren nucleus is *now* planar. Thus both the forms VI



IX

and VII would give the same tetrabromide, IX (the same is true for the tetrahydroxide); the tetrabromide should thus be planar, the configuration of each arsenic atom being trigonal bipyramidal in the 5-covalent state (Fig. 2).

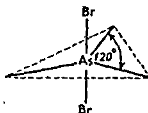


FIG. 6.2.

Quinquevalent phosphorus and arsenic can make use of the *3d* or *4d* orbitals, respectively (*cf.* nitrogen, §2b). Thus nitrogen has a maximum covalency of 4, whereas that of phosphorus and arsenic is 5 or 6, *e.g.*, the covalency of 6 is exhibited by phosphorus in *solid* phosphorus pentachloride; X-ray diffraction shows this "molecule" (in the solid state) is $PCl_4^+ PCl_6^-$.

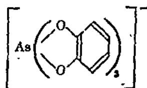
Phosphorus, which is $(1s^2)(2s^2)(2p^6)(3s^2)(3p^3)$ in the ground state, may become $(1s^2)(2s^2)(2p^6)(3s)(3p^3)(3d)$ in its "valence state", since the 3s and 3d orbitals have energy levels which are close together.

the corners of an equilateral triangle, and two orbitals perpendicular to this plane (see Fig. 2). It so happens that spd^3 hybridisation would also give a trigonal bipyramid; it is not certain which hybridisation is actually present in quinquevalent phosphorus. Electron diffraction studies of the vapours of phosphorus pentachloride and pentafluoride indicate the trigonal bipyramidal configuration in these molecules. The phosphonium ion might possibly be formed from this trigonal bipyramid by the transference of one

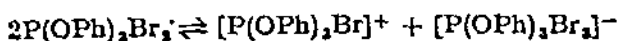
of the electrons, or by the transference of a 3*s* electron and hybridisation of the $(3s)(3p^3)$ orbitals; in either case the tetrahedral configuration of the phosphonium ion can be asymmetric, but only in the case of the hybridisation of the $(3s)(3p^3)$ orbitals will the four bonds be equivalent. Since the properties of phosphonium compounds are in agreement with the equivalence of the four bonds, it therefore appears, on theoretical grounds, that the tetrahedral configuration with the phosphorus atom at the centre is the probable one.

From the experimental side, the preparation of optically active spiro-compounds of phosphorus (§3b) and of arsenic (§4a) proves the tetrahedral configuration of these atoms. Earlier work by Mann *et al.* (1936, 1937) has also definitely established this configuration. These authors prepared compounds of the type $[R_3As—CuI]_4$ by combination of tertiary arsines or phosphines with cuprous iodide (or silver iodide); in these compounds the phosphorus or arsenic is 4-covalent, and X-ray analysis studies of the arsenic compound showed that the arsenic atom is at the centre of a tetrahedron. Since the corresponding phosphorus compounds are isomorphous, the configuration of the phosphorus is also tetrahedral.

In the *solid* state, phosphorus and arsenic compounds may contain a negatively charged phosphorus or arsenic atom, *e.g.*, PCl_4^+ PCl_6^- (see above). In this condition, the phosphorus acquires an electron to become ---$(3s)(3p^3)(3d^2)$, and the arsenic also acquires an electron to become ---$(4s)(4p^3)(4d^2)$. In both cases the configuration is octahedral (six sp^3d^2 bonds), *e.g.*, the following compound has been resolved (Rosenheim *et al.*, 1925).



Harris *et al.* (1956) have shown that a negatively charged phosphorus atom can also exist in *solution*; these authors showed that triphenyl phosphite dibromide ionises in methyl cyanide solution as follows:

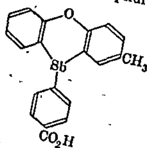


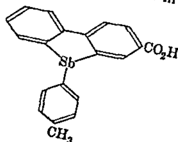
$$2P(OPh)_3Br_2 \rightleftharpoons [P(OPh)_3Br]^+ + [P(OPh)_3Br_3]^-$$

§4c. Stereochemistry of antimony compounds. Two optically active tervalent antimony compounds have been prepared, the phenoxstibine (I) and the stibiafluorene (II; Campbell, 1947, 1950). The asymmetry in I is probably due to ... about the O—Sb axis (*cf.* phenoxarsines, §4b).

It is of interest to note, in ... ction, ... s by

Weston (1954) have led him to the conclusion that tervalent antimony, arsenic and sulphur compounds should be stable to inversion



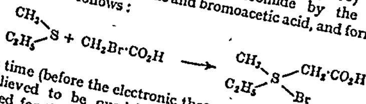


at room temperature. On the other hand, similar compounds of phosphorus would be optically stable only at low temperatures, and those of nitrogen not at all.

§5. STEREOCHEMISTRY OF SULPHUR COMPOUNDS

Various types of sulphur compounds have been obtained in optically active forms, and although the general picture of the configurations of these molecules is quite clear, the details of the nature of the bonds of the central sulphur atom are in a state of flux. In the following account the "octet shell" for sulphur will be used, but an alternative explanation is offered in §5e.

§5a. **Sulphonium salts.** Pope and Peachey (1900) prepared carboxymethylethylmethylsulphonium bromide by the reaction between ethyl methyl sulphide and bromoacetic acid, and formulated the reaction as follows:



At this time (before the electronic theory of valency, 1916), sulphur was believed to be quadricovalent, and so Pope and Peachey accounted for the optical activity of this compound (see below) by assuming that the sulphur atom was at the centre of a tetrahedron, *i.e.*, the configuration was similar to carbon. According to the electronic theory of valency, however, sulphur is tercovalent unielectrovalent in sulphonium salts, and so the sulphur is believed to be at the centre of a tetrahedron with three corners occupied by the groups CH_3, C_2H_5 and $CH_2{\cdot}CO_2H$, and the fourth corner

by a lone pair of electrons (see Fig. 3). This molecule is not superimposable on its mirror image, and hence can, at least theoretically, exist in two optically active forms. This bromide was treated with silver (+)-camphorsulphonate and the salt obtained was

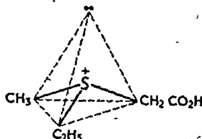
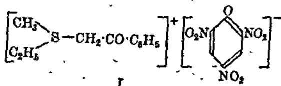
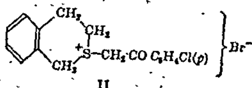


FIG. 6 3.

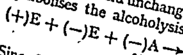
fractionally crystallised from a mixture of ethanol and ether. Pope and Peachey found that the (+)-sulphonium camphorsulphonate was the less soluble fraction, and had an M_D of +68°. Since the rotation of the (+)-camphorsulphonate ion is about +52°, this leaves +16° as the contribution of the sulphonium ion to the total rotation (see §12. I). Although this does not prove conclusively that the sulphur compound is optically active, it is certainly strong evidence in its favour. Final proof was obtained by replacement of the camphorsulphonate ion by the platinichloride ion to give $[CH_3(C_2H_5)\cdot S\cdot CH_2\cdot CO_2H]_2{}^+PtCl_6{}^=$; this compound had an $[\alpha]_D$ of +4·5° in water. In a similar way, Smiles (1900) prepared ethylmethylphenacylsulphonium picrate, I, in two optically active



forms, one with an $[\alpha]_D$ of +8·1° and the other −9·2°. A more recent example of an optically active sulphonium salt is one with the sulphur atom in a ring; this compound, II, was obtained as the optically active ion with the picrate (Mann and Holliman, 1946).



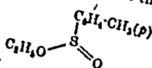
§5b. Sulphinic esters. Phillips (1925) partially resolved sulphinic esters, $R{\cdot}SO_2R'$, by means of the kinetic method of resolution (§10 vii. II). Two molecules of ethyl *p*-toluenesulphinate were heated with one molecule of (−)-menthyl alcohol or (−)-*sec.*-octyl alcohol, *i.e.*, the sulphinate was subjected to alcoholysis. Now, if the sulphinate is a racemic modification, then the (+)- and (−)-forms will react at *different* rates with the optically active alcohol (see §§2, 7b. II). Phillips actually found that the (+)-ester reacted faster than the (−)-ester. If we represent the ester by E, the alcohol by A, and unchanged ester by E_r, then the following equation symbolises the alcoholysis:



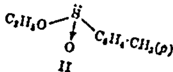
$$(+)E + (-)E + (-)A \rightarrow [(+)E(-)A] + [(-)E(-)A] + (+)E_r + (-)E_r$$

Since [(+)E(−)A] is greater than [(−)E(−)A], it therefore follows that $(+)E_r$ is less than $(-)E_r$; thus a partial resolution has occurred. The unchanged ester, having a lower boiling point than the new ester, distilled off first; this contained more of the (−)-form. The residual ester (the higher boiling fraction) was then heated with a large excess of ethanol; alcoholysis again occurred, this time the (−)-alcohol (menthol or octyl) being displaced to regenerate the original ethyl *p*-toluenesulphinate. This resulted in a fraction containing more of the (+)-form.

To account for the optical activity of these sulphinates, the older formula I, with quadricovalent sulphur linked to the oxygen atom by a double bond, was replaced by formula II, in which the sulphur atom is at the centre of the tetrahedron, but one corner is occupied



I

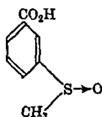


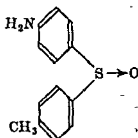
II

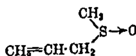
by a lone pair of electrons (*cf.* Fig. 3). In I, the sulphur atom is at the centre of a tetrahedron, and the molecule is *flat*, and consequently is superimposable on its mirror image. Molecule II, however, is asymmetric, and so is optically active.

§5c. Sulphoxides. Sulphoxides of the type $R{\cdot}SO{\cdot}R'$ have also been resolved, and so the formula $RR'S{=}O$ was rejected and replaced by $RR'S\rightarrow O$; sulphoxides I and II were resolved by Phillips *et al.* (1926), and Karrer *et al.* (1931) obtained III in the (−)-form and the racemic modification.

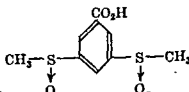
Bell and Bennett (1927) investigated disulphoxides of the type



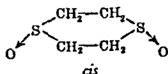




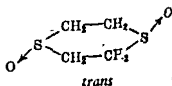
$CH_3{\cdot}SO{\cdot}CH_2{\cdot}CH_2{\cdot}SO{\cdot}CH_3$. This molecule contains two similar asymmetric carbon atoms and so is of the type *Cabd·Cabd*. Thus it should exist in one racemic modification and one *meso*-form. Bell and Bennett failed to resolve this compound, but succeeded in resolving the following disulphoxide.



If the former disulphoxide (the dioxide of a 1:4-dithian) is converted into the corresponding ring compound (*i.e.*, into a cyclic 1:4-dithian), then two geometrical isomers are possible, neither of which is resolvable; these two forms have been isolated by Bell and Bennett (1927, 1929).

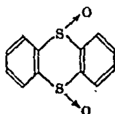


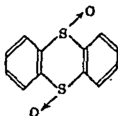
cis



trans

Thianthren dioxide also exists in two geometrical isomeric forms. The structure is probably folded (as is that of *thianthren*), but only two forms would be possible, since in the third the two oxygen



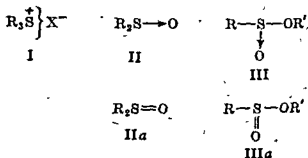


atoms would be coincident (*cf.* the arsanthrens, §4b). The two forms are the *cis* and *trans*, and have been distinguished by means

§5e. The valency disposition of the sulphur atom. The electronic configuration of sulphur is $(1s^2)(2s^2)(2p^6)(3s^2)(3p^4)$; the four $3p$ orbitals are $(3p_x{}^2)(3p_y)(3p_z)$.

Bivalent sulphur. In compounds of the type Sa_2, sulphur uses the two unshared $3p$ orbitals, $3p_y$ and $3p_z$, and therefore the valency angle should be in the region of 90°. In hydrogen sulphide the valency angle has been found to be approximately 92°, and in sulphur dichloride approximately 103°; the large value in the latter case is partly due to the repulsion between the chlorine atoms. It has also been suggested that pure p-bonds are not present in these compounds, but that the bonds are admixtures of s and p orbitals, and possibly d orbitals (see below). The variation of the bond angle with the nature of the groups attached is thus attributed partly to changes in admixture of these orbitals.

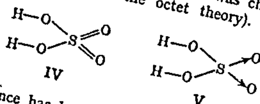
Tercovalent and quadricovalent sulphur. Sulphur could become tercovalent unielectrovalent by the transfer of one of its $3p_x$ electrons, *e.g.*, in sulphonium salts, I. Sulphur could also become tercovalent by *donating* the lone $3p_x$ pair of electrons, *e.g.*, in sulphoxides, II, and in sulphinic esters, III. On the other hand, sulphur could be quadricovalent in sulphoxides (II*a*) and in



sulphinic esters (III*a*), but in this case, if the four are arranged tetrahedrally with the sulphur at the centre (as thought originally), then the molecules would be flat and hence not resolvable. Since, however, they are resolvable, the sulphur atom cannot be in the same plane as the other three groups. This therefore led Phillips to propose structures II and III, and according to him the arrangement was tetrahedral, with one corner unoccupied (see Fig. 3). Further evidence for the non-planar configuration of sulphoxides and sulphinic esters is given by the fact that in thionyl chloride the Cl—S—Cl angle is approximately 114°, and the O—S—Cl angle approximately 106°. These values are in agreement with a tetrahedral configuration with the sulphur atom at the apex of the tetrahedron. Structures II and III are therefore satisfactory, but

there are, however, alternative electronic structures which fit the facts equally well (see below).

Sexavalent sulphur. X-ray analysis of the sulphate ion shows that it has a tetrahedral configuration; and sulphuryl chloride also has been shown to be tetrahedral. Originally, sulphur was assumed to be sexavalent in these compounds, and so the formula of sulphuric acid was written as IV, but when the electronic theory of valency was introduced, this formula was changed to V (to satisfy the requirements of the octet theory). More recently,



however, evidence has been obtained (mainly from bond-length measurements) that sulphur is sexavalent in sulphur hexafluoride, sulphur trioxide, sulphuric acid, etc., *i.e.*, sulphuric acid is IV and *not* V. Now, if sulphur is sexavalent in sulphur hexafluoride, it might be anticipated that all six fluorine atoms are equivalent. This has actually been found to be so, and it therefore follows that all the six sulphur valencies are equivalent. A configuration which fits these facts is an octahedral one (Fig. 4 *a*). In sexavalent compounds, the electronic configuration of sulphur would be $(1s^2)(2s^2)(2p^6)(3s)(3p^3)(3d^2)$, which is brought about by the pro motion of one $3s$ and one $3p$ electron to $3d$; this is possible becaus of the small energy differences between the orbitals concerned. Hybridisation of the $(3s)(3p^3)(3d^2)$ electrons would give six equivalent bonds (six sp^3d^2 bonds) arranged octahedrally (Fig. 4*a*).

As we have seen above, thionyl chloride has a tetrahedral configuration, and the sulphoxides and sulphinic esters, since they are

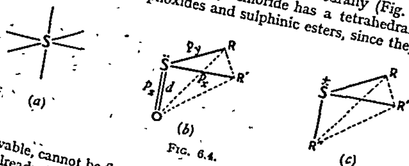


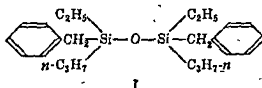
FIG. 6.4.

resolvable, cannot be flat. The optical activity of these compounds has already been explained on the basis of the octet theory, the sulphur atom being linked to the oxygen atom by a dative bond

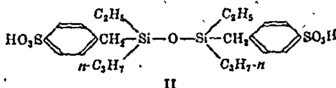
and one corner of the tetrahedron being unoccupied (except by a lone pair of electrons). The optical activity of these compounds, however, can also be explained on the assumption that the oxygen atom is linked to the sulphur atom by a double bond. In 4-covalent sulphur compounds, the electronic configuration of the sulphur atom could be $(1s^2)(2s^2)(2p^6)(3s^2)(3p^3)(3d)$, which is brought about by the promotion of one $3p$ electron to $3d$. In this arrangement, the two R—S bonds are formed by two of the three $3p$ electrons of the sulphur atom, and the S=O bond by one $3p$ electron and the $3d$ electron of the sulphur atom (this is a π-bond of the type p_π—d_π; in ethylene the π-bond is of the type p_π—p_π). The configuration of the molecule in this arrangement is still tetrahedral, with the sulphur atom at one corner of the tetrahedron (Fig. 4 *b*). This valency configuration is supported by bond length measurements of the S—O bonds in sulphoxides and sulphones; in both the bonds are double bonds.

In the sulphonium salts, the transfer of one of the $3p$ electrons of the sulphur atom (from the ground state) leaves the sulphur atom in a tercovalent unielectrovalent state (see above). Thus, in sulphonium salts, the sulphur atom is again at the corner of a tetrahedron (Fig. 4 *c*).

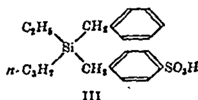
§6. Stereochemistry of silicon compounds. Kipping (1907) prepared benzylethylpropylsilicyl oxide, I, and isolated one *form* of it. If the silicon atom has a tetrahedral configuration, this



I



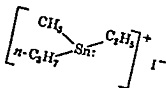
II



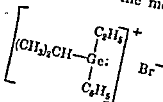
III

molecule is of the type $Cabd{\cdot}Cabd$, *i.e.*, it should exist in (+)-, (−)-, and *meso*-forms. When I was sulphonated to give II, the latter compound was resolved. Challenger and Kipping (1910) also resolved the silicon compound III.

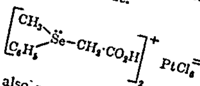
§7. Stereochemistry of tin compounds. Pope and Peachey (1900) obtained ethylmethyl-*n*-propylstannonium iodide in the dextrorotatory form; concentration of the mother liquor also gave this (+)-form. Thus we have an example of asymmetric transformation (§10 iv. II).



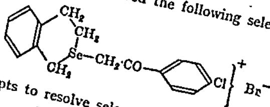
§8. Stereochemistry of germanium compounds. Schwarz and Lewinsohn (1931) obtained the (+)-form of ethylphenylisopropylgermanium bromide, but failed to get the (−)-form; this latter form appears to racemise in the mother liquor.



§9. Stereochemistry of selenium compounds. Pope *et al.* (1902) resolved carboxymethylmethylphenylselenonium bromide in the same way as the corresponding sulphonium salts (§5a); they obtained the active platinichloride.

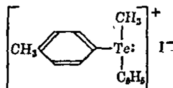


Mann *et al.* (1945) also resolved the following selenonium salt:



So far, attempts to resolve selenoxides have failed.

§10. Stereochemistry of tellurium compounds. Lowry *et al.* (1929) obtained the optically active forms of methylphenyl-*p*-tolyltelluronium iodide.



READING REFERENCES

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). Ch. 4, pp. 400–443. Optical Isomerism of Elements Other Than Carbon

Wheland, *Advanced Organic Chemistry*, Wiley (1949). (i) Ch. 8. The Configuration and Stereochemistry of Elements Other Than Carbon. (ii) Ch. 12. Molecular Rearrangements. The 1,2-Shifts.

Sidgwick, *Organic Chemistry of Nitrogen*, Oxford Press (1942, 2nd ed. by Taylor and Baker). Ch. 6. Hydroxylamine Derivatives.

Blatt, The Beckmann Rearrangement, *Chem. Reviews*, 1933, **12**, 215.

Brynmor Jones, Kinetics and Mechanism of the Beckman Rearrangement, *Chem. Reviews*, 1944, **35**, 335.

Dewar, *The Electronic Theory of Organic Chemistry*, Oxford Press (1949). Ch. 10, p. 219. The Beckmann Rearrangement.

Hornung and Stromberg, Beckmann Rearrangements. Aldoximes, *J. Amer. Chem. Soc.*, 1952, **74**, 5151.

Mann, Some Aspects of the Organic Chemistry of Phosphorus and Arsenic, *J.C.S.*, 1945, **65**.

Mann, *The Heterocyclic Derivatives of P, As, Sb, Bi, and Si*, Interscience Publishers (1950).

Stephen and Staskun, A New Mechanism for the Beckmann Rearrangement of Oximes, *J.C.S.*, 1956, 980.

Abrahams, The Stereochemistry of Sub-group VIB of the Periodic Table, *Quart. Reviews (Chem. Soc.)*, 1956, **10**, 407.

McCasland and Proskow, Synthesis of an Image-Superposable Molecule which Contains no Plane or Centre of Symmetry, *J. Amer. Chem. Soc.*, 1956, **78**, 5646.

Klyne and de la Mare (Ed.), *Progress in Stereochemistry*, Butterworth. Vol. II (1958). Ch. 6. The Stereochemistry of the Group V Elements.

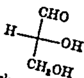
CHAPTER VII

CARBOHYDRATES

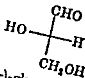
This chapter is mainly concerned with the stereochemistry of the carbohydrates and the structures of the disaccharides and polysaccharides. It is assumed that the reader is familiar with the open-chain structures and general reactions of the monosaccharides (for an elementary account of these compounds, see Vol. I, Ch. XVIII).

§1. DETERMINATION OF THE CONFIGURATION OF THE MONOSACCHARIDES

Aldotrioses. There is only one aldotriose, and that is glyceraldehyde. As we have seen (§5. II), the enantiomorphs of this compound have been chosen as the *arbitrary* standards for the D- and L-series in sugar chemistry:

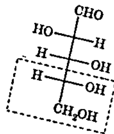


D(+)-glyceraldehyde



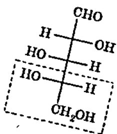
L(−)-glyceraldehyde

The conventional planar diagrams of the sugars are always drawn with the CHO (or $CH_2OH{\cdot}CO$) group at the top and the CH_2OH group at the bottom; the following short-hand notation is also used:





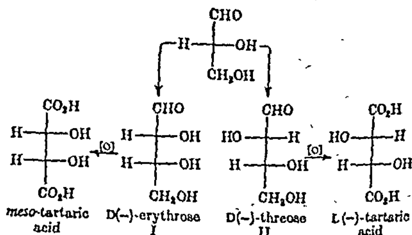
D-series





L-series

Aldotetroses. The structural formula of the aldotetroses is $CH_2OH{\cdot}CHOH{\cdot}CHOH{\cdot}CHO$. Since this contains two unlike

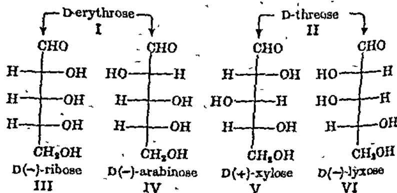


asymmetric carbon atoms, there are four optically active forms (two pairs of enantiomorphs) possible theoretically. All four are known, and correspond to D- and L-threose and D- and L-erythrose. D(+)-Glyceraldehyde may be stepped up by the Kiliani reaction to give D(−)-erythrose and D(−)-threose. The question now is: Which is which? On oxidation, D-erythrose gives *meso*tartaric, and on reduction gives *meso*erythritol. Therefore D-erythrose is I, and consequently II must be D-threose. The configuration of the latter is confirmed by the fact that on oxidation, D-threose gives L(−)-tartaric acid.

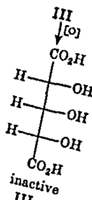
Aldopentoses. These have the structural formula


$$CHO \cdot CHOH \cdot CHOH \cdot CHOH \cdot CH_2OH,$$

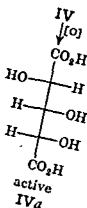
and since it contains three unlike asymmetric carbon atoms, there are eight optically active forms (four pairs of enantiomorphs). All are known, and correspond to the D and L forms of ribose, arabinose, xylose and lyxose. Their configurations may be ascertained by either of the following two methods.



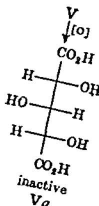
One method starts by stepping up the aldotetroses by the Kiliani reaction. Thus D-erythrose gives D(−)-ribose and D(−)-arabinose; similarly, D-threose gives D(+)-xylose and D(−)-lyxose. III and IV must be ribose and arabinose, but which is which? On oxidation with nitric acid, arabinose gives an optically active dicarboxylic acid (a trihydroxyglutaric acid), whereas ribose gives an optically inactive dicarboxylic acid. When the terminal groups, *i.e.*, CHO and CH_2OH, of III are oxidised to carboxyl groups, the molecule produced (III*a*) possesses a plane of symmetry, and so is inactive. Oxidation of IV gives IV*a*, and since this molecule has no plane (or any other element) of symmetry, it is optically active. Thus III is D-ribose and IV is D-arabinose.



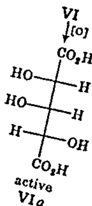
inactive
III*a*



active
IV*a*



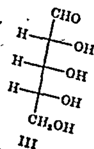
inactive
V*a*



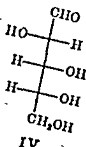
active
VI*a*

V and VI must be xylose and lyxose, but which is which? The former sugar, on oxidation, gives an optically inactive dicarboxylic acid, whereas the latter gives an optically active dicarboxylic acid. Therefore V is D-xylose and VI is D-lyxose.

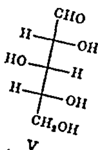
The following is the alternative method of elucidating the configurations of the aldopentoses; it is more in keeping with Fischer's solution to the problem. The structural formula of the aldopentoses can give rise to four pairs of enantiomorphs, the D-forms of which are as follows:



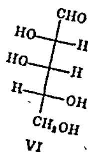
III



IV

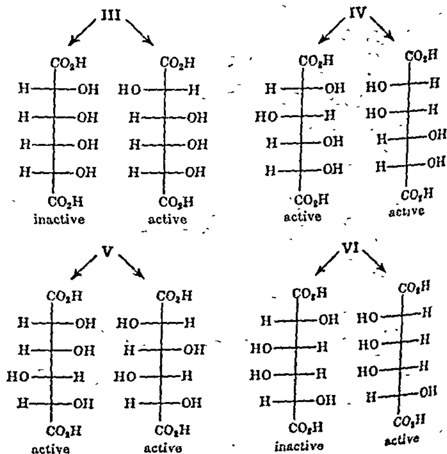


V



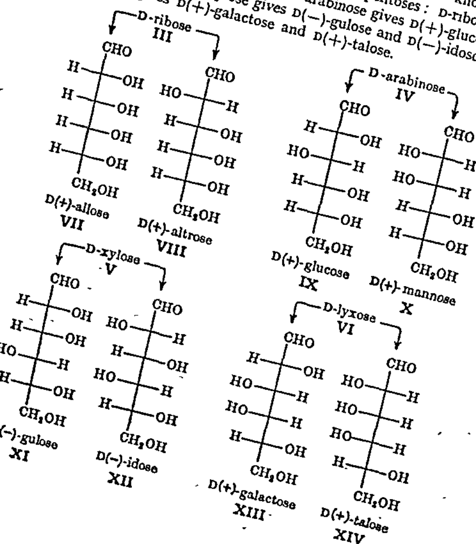
VI

It should be noted that these four configurations have been obtained from first principles (see §7c. II); no recourse has been made to the configurations of the aldotetroses. Arabinose and lyxose, on oxidation with nitric acid, produce optically active dicarboxylic acids (trihydroxyglutaric acids). Therefore these two pentoses must be IV and VI, but we cannot say which is which. Xylose and ribose, on oxidation, produce optically inactive dicarboxylic acids (trihydroxyglutaric acids). Therefore these two pentoses must be III and V, and again we cannot say which is which. When each aldopentose is stepped up by one carbon atom (by means of the Kiliani reaction) and then oxidised to the dicarboxylic acid (the terminal groups are oxidised), it is found that arabinose and xylose each give *two* active dicarboxylic acids, whereas ribose and lyxose each give one active and one inactive (*meso*) dicarboxylic acid. The following chart shows the dicarboxylic acids obtained from the configurations III–VI.



It therefore follows that D-ribose is III, D-arabinose is IV, D-xylose is V, and D-lyxose is VI. These configurations are confirmed by the facts that ribose and arabinose give the same osazone, and xylose and lyxose give the same osazone; the only difference between sugars giving the same osazone is the configuration of the second carbon atom, *i.e.* III and IV are epimers, as are V and VI. It should also be noted that arabinose and lyxose produce the same trihydroxyglutaric acid on oxidation.

Aldohexoses. The structural formula of these compounds is CHO·CHOH·CHOH·CHOH·CHOH·CH_2OH, and since it contains four unlike asymmetric carbon atoms, there are sixteen optically active forms (eight pairs of enantiomorphs). All are known, and may be prepared by stepping up the aldopentoses: D-ribose gives D(+)-allose and D(+)-altrose; D-arabinose gives D(+)-glucose and D(+)-mannose; D-xylose gives D(−)-gulose and D(−)-idose; and D-lyxose gives D(+)-galactose and D(+)-talose.



VII and VIII must be allose and altrose, but which is which? On oxidation with nitric acid, the former gives an optically inactive (allomucic) and the latter an optically active (talomucic) dicarboxylic acid. Therefore allose is VII and altrose is VIII.

XIII and XIV must be galactose and talose, but which is which? On oxidation with nitric acid, the former gives an optically inactive (mucic) and the latter an optically active (talomucic) dicarboxylic acid. Therefore XIII is galactose and XIV is talose.

The elucidation of the configurations of the remaining four aldohexoses is not quite so simple, since, on oxidation with nitric acid, glucose and mannose *both* give optically active dicarboxylic acids, as also do gulose and idose; in all four configurations (IX, X, XI, XII), replacement of the two terminal groups (CHO and CH_2OH) by carboxyl groups leads to dicarboxylic acids whose structures have no plane (or any other element) of symmetry. It has been found, however, that the dicarboxylic acid from glucose (saccharic acid) is the same as that obtained from gulose (actually the two saccharic acids obtained are enantiomorphous, D-glucose giving D-saccharic acid and D-gulose L-saccharic acid). Since saccharic acid, $CO_2H \cdot (CHOH)_4 \cdot CO_2H$, is produced by the oxidation of the terminal groups with the rest of the molecule unaffected, it therefore follows that the "rest of molecule" must be the same for both glucose and gulose. Inspection of formulae IX, X, XI and XII shows that only IX and XI have the "rest of the molecule" the same; by interchanging the CHO and CH_2OH groups of IX, the enantiomorph of XI, *i.e.*, L-gulose, is obtained. Therefore IX must be glucose (since we know that glucose is obtained from arabinose), and XI must be gulose. Consequently X is mannose and XII is idose.

Ketohexoses. All the ketohexoses that occur naturally have the ketonic group adjacent to a terminal CH_2OH group, *i.e.*, the structural formula of all the natural ketohexoses is


$$CH_2OH \cdot CO \cdot CHOH \cdot CHOH \cdot CHOH \cdot CH_2OH.$$

Since this structure contains three dissimilar asymmetric carbon atoms, there are eight optically active forms (four pairs of enantiomorphs) possible theoretically; of these the following six are known: D(−)- and L(+)-fructose, D(+)- and L(−)-sorbose, D(+)-tagatose and L(−)-psicose. Only D(−)-fructose, L(−)-sorbose and D(+)-tagatose occur naturally.

Fructose. Natural fructose is lævorotatory, and since D-glucose gives the *same* osazone as natural fructose, the latter must be D(−)-fructose. Furthermore, since osazone formation involves only the first two carbon atoms in a sugar, it therefore follows that

the configuration of the rest of the molecule in glucose and fructose must be the same. Hence the configuration of D(−)-fructose is XV, and is confirmed by the fact that D(+)-glucose may be converted into D(−)-fructose *via* the osazone (see following chart).

CH_2OH / CO / HO—C—H / H—C—OH / H—C—OH / CH_2OH
D(−)-fructose
XV

$\xrightarrow{C_6H_5NHNH_2}$

CH=N·NH·C_6H_5 / C=N·NH·C_6H_5 / HO—C—H / H—C—OH / H—C—OH / CH_2OH
osazone

$\xleftarrow{C_6H_5NHNH_2}$

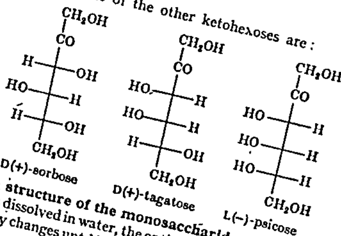
CHO / H—C—OH / HO—C—H / H—C—OH / H—C—OH / CH_2OH
D(+)-glucose

osazone $\xrightarrow{\text{hydrolysis}}$ osone

CHO / CO / HO—C—H / H—C—OH / H—C—OH / CH_2OH
osone

osone $\xrightarrow{Zn/CH_3CO_2H}$ D(−)-fructose

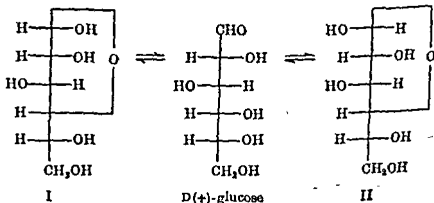
The configurations of the other ketohexoses are:



§2. Ring structure of the monosaccharides. When a monosaccharide is dissolved in water, the optical rotatory power of the solution gradually changes until it reaches a constant value (Dubrunfaut,

1846), *e.g.*, a freshly prepared solution of glucose has a specific rotation of +111°, and when this solution is allowed to stand, the rotation falls to +52·5°, and remains constant at this value. The final stage can be reached *more rapidly either* by heating the solution or by adding some catalyst which may be an acid or a base. This change in specific rotation is known as **mutarotation**; all *reducing* sugars (except a few ketoses) undergo mutarotation.

To account for mutarotation, Tollens (1883) suggested an oxide ring structure for D(+)-glucose, whereby *two* forms would be produced, since, in the formation of the ring, another asymmetric carbon atom (which can exist in *two* configurations) is produced (*cf.* the Kiliani reaction). Tollens assumed that a five-membered ring (the γ-form) was produced:



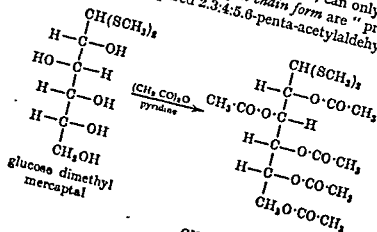
The difficulty of this suggestion was that there was no experimental evidence for the existence of these two forms. Tanret (1895), however, isolated two isomeric forms of D(+)-glucose, thus apparently verifying Tollens' supposition (but see §§7a, 7f). *The two forms,* I *and* II, *are known respectively as* α- *and* β-D(+)-γ-*glucose* (see also §7b for the nomenclature of these forms).

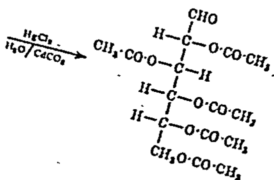
Ring formation of a sugar is really hemiacetal formation, one alcoholic group of the sugar forming a hemiacetal with the aldehyde group of the *same* molecule, thus producing a ring structure which is known as the *lactol* form of the sugar.

Mechanism of mutarotation. According to Lowry (1925), *mutarotation is not possible without the presence of an amphiprotic* solvent, *i.e.*, a solvent which can function both as an acid and a base, *e.g.*, water. Thus Lowry and Faulkner (1925) showed that mutarotation is arrested in pyridine solution (basic solvent) and in cresol solution (acidic solvent), but that it takes place in a *mixture of pyridine and cresol.* It has been assumed that when mutarotation takes place, the ring opens and then recloses in the inverted

position or in the original position. There is some evidence for the existence of this open-chain form. The absorption spectra of fructose and sorbose in aqueous solution indicate the presence of open-chain forms; aldoses gave negative results (Bednarczyk *et al.*, 1938). Solutions of glucose and arabinose in 50 per cent. sulphuric acid gave an ultraviolet absorption spectrum containing the band characteristic of the oxo (carbonyl) group (Pascu *et al.*, 1948). Aldoses in solution contain a form which is reducible at the dropping mercury electrode (Cantor *et al.*, 1940). Although the nature of this reducible form has not been established, it is probably the open-chain form, either free or hydrated. Furthermore, a relationship was shown to exist between the amount of this reducible form and the rate of mutarotation. One interpretation of this observation is that the reducible form is an intermediate in mutarotation. Rate constants for the conversion of the ring forms of aldoses to the open-chain form have been calculated from polarographic measurements, and it has also been shown that the energy of activation required to open the pyranose ring is the same for glucose, mannose, galactose, arabinose and xylose (Delahay *et al.*, 1952).

It is interesting to note in connection with this problem of the existence of the open-chain structure, that *aldehydo-sugars, i.e.*, aldoses in which the aldehyde group is present, can only be *isolated* if all the hydroxyl groups in the *open-chain form* are "protected"; *e.g.*, Wolfrom (1929) prepared 2.3:4:5.6-penta-acetylaldehydoglucose as follows:





The problem now is: *What is the mechanism of the formation of the open-chain form from the ring-form?* Lowry (1925) suggested that it occurred by the simultaneous addition and elimination of a proton, since both an acid and a base must be present (see above). This concerted mechanism would conform to a third-order reaction:



Swain *et al.* (1952) have shown that the mutarotation of tetramethylglucose, catalysed by phenol and pyridine in benzene solution, is a third-order reaction; this supports the above mechanism. On the other hand, some authors believe that the reaction proceeds in two independent ways, one being an acid-catalysed reaction, and the other a base-catalysed reaction. In this case the mechanism would conform to a second-order reaction. Hill *et al.* (1952) have shown that the mutarotation of glucose in aqueous methanol containing acetate buffers is in better agreement with a second-order reaction than with a third-order.

It can thus be seen that the [illegible] be regarded as settled, and it a[illegible] gated (free or as a derivative) a[illegible] play a part in deciding which mechanism will operate.

Preparation of the α- and β-forms of a sugar. Experimentally, it is very difficult to isolate the α- and β-forms of a sugar. The ordinary form of D(+)-glucose is the α-isomer, m.p. 146° and $[\alpha]_D = +111°$; this form may be prepared by crystallising glucose from cold ethanol. The β-isomer, m.p. 148–150°, $[\alpha]_D = +19.2°$, can be obtained by crystallising glucose from hot pyridine. Thus the α-form may be converted into the β-, and *vice versa*, during the process of crystallisation; this is an example of asymmetric transformation (§10 iv. II). Both forms show mutarotation, the final value of the specific rotation being +52.5°; this corresponds to a mixture containing about 38 per cent. of the α-isomer, and 62 per cent. of the β-. The two stereoisomeric ring forms of a sugar are often referred to as *anomers*.

Summary of the evidence for the ring structures of sugars. The cyclic structure of the sugars accounts for the following facts:

(i) The existence of two isomeric forms (anomers) of a given sugar, *e.g.*, α- and β-glucose.

(ii) Mutarotation.

(iii) Glucose and other aldoses do not give certain characteristic reactions of aldehydes, *e.g.*, Schiff's reaction, do not form a bisulphite or an aldehyde-ammonia compound. Recently, however, it has been shown that by preparing Schiff's reagent in a special way, it becomes very sensitive, simple aldoses restoring the pink colour to this solution; the monosaccharide aldoses react strongly, but the disaccharide aldoses react weakly (Tobie, 1942). This reaction with a sensitive Schiff's reagent appears to indicate that some, although a very small amount, of the open-chain form of a sugar is present in solution in equilibrium with the two ring forms.

(iv) Glucose penta-acetate does not react with hydroxylamine; this indicates that the aldehyde group is absent in this derivative (glucose itself does form an oxime).

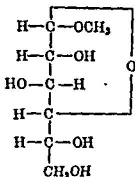
(v) Aldehydes normally form acetals by combination with two molecules of a monohydric alcohol; aldoses (and ketoses) combine with only one molecule of an alcohol. It should be noted, however, that aldoses will combine with *two* molecules of a thiol to form a mercaptal (thioacetal).

(vi) X-ray analysis definitely proves the existence of the ring structure, and at the same time indicates the size of the ring (see §7f).

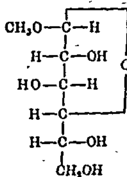
§3. Glycosides. Just as simple hemi-acetals react with another molecule of an alcohol to form acetals, so can the sugars, in their ring forms (lactols), react with a molecule of an alcohol to form the acetal derivative, which is known under the generic name of **glycoside**; those of glucose are known as *glucosides*; of fructose, *fructosides*, etc. The hydroxyl group produced at the oxo group by ring formation is known as the *glycosidic hydroxyl group*. This group can be acetylated and methylated, as can all the other hydroxyl groups in the sugar, but the glycoside derivatives are far more readily decomposed by various reagents.

E. Fischer (1893) refluxed glucose in methanol solution in the presence of one half per cent. hydrochloric acid, and thereby obtained a white crystalline product which contained *one* methyl group (as shown by analysis), and which did not reduce Fehling's solution or mutarotate, and did not form an osazone. Thus the hemiacetal structure is no longer present in this compound; in fact, this compound appears to be an acetal since it is stable in alkaline solution (Fehling's solution). Furthermore, on boiling with dilute inorganic acids, the compound regenerated the original sugar, a reaction again typical of acetals. Ekenstein (1894) isolated a second isomer from the reaction mixture when he repeated Fischer's work, and Fischer

explained the existence of these two isomers by suggesting ring structures for the two methyl glucosides, *viz.*,



methyl α-D-glucoside



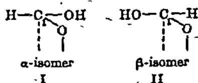
methyl β-D-glucoside

Fischer assumed that these methyl glucosides were five-membered ring systems, basing his assumption on Tollens' suggestion (§2). As we shall see later (§7a), Fischer's assumption is incorrect.

The non-sugar part of a glycoside is known as the *aglycon* (or *aglycone*), and in many glycosides that occur naturally, the aglycon is often a phenolic compound (see §24).

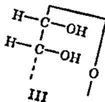
Fischer (1894) found that methyl α-D-glucoside was hydrolysed by the enzyme maltase, and the β-D-glucoside by the enzyme emulsin. Furthermore, Fischer also found that maltase would not hydrolyse the β-glucoside, and that emulsion would not hydrolyse the α-glucoside. Thus the two isomers can be distinguished by the specificity of action of certain enzymes (see also §16. XIII). Armstrong (1903) followed these enzymic hydrolyses polarimetrically, and showed that methyl α-D-glucoside liberates α-D-glucose, and that the β-glucoside liberates β-D-glucose; Armstrong found that hydrolysis of the α-glucoside produced a "downward" mutarotation, whereas that of the β-glucoside produced an "upward" mutarotation. It therefore follows that α-D-glucose is stereochemically related to methyl α-D-glucoside, and β-D-glucose to methyl β-D-glucoside.

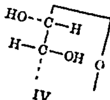
§4. Configuration of C_1 in glucose. The configurations of C_1 in α- and β-D-glucose have been written, in the foregoing account, as:



The question that now confronts us is: What justification is there for this choice, *i.e.*, what is the evidence that enables us to say that the α-isomer (characterised by certain physical constants) actually has the hydrogen atom to the left and the hydroxyl group to the right? Hudson (1909) proposed the *empirical* rule that of an α, β pair of sugars in the D-series, the α-isomer, which has the *higher* dextrorotation (*i.e.*, this physical constant decides which of the two is to be designated α-), has the hydrogen to the left (*i.e.*, I); the β-isomer consequently has the hydrogen atom to the right (II). Thus α-D(+)-glucose is the isomer with the specific rotation +111°, and β-D(+)-glucose is the isomer with the specific rotation +19·2°. If the D-sugar has a negative rotation, then according to the empirical rule, the β-isomer has the higher negative rotation (*i.e.*, the less positive rotation), *e.g.*, α-D(−)-fructose is the isomer with the specific rotation −20°, and the β-isomer −133°. In the L-sugars, the α-isomer is the one with the *higher* lævorotation, and the other is the β-isomer; thus the α-forms (and the β-forms) of the D- and L-series are enantiomorphous.

Böeseken (1913) found that when boric acid is added to a solution of a cyclic 1:2-glycol, the electrical conductivity of the solution is greater than that of boric acid itself, and that the increase is greater for the *cis*-isomer than for the *trans*- (see Vol. I). This phenomenon has been used to distinguish between the two anomers of D-glucose; the results obtained showed that the conductivity of the isomer called the α (from the above empirical rule), in the presence of boric acid, decreased during mutarotation, whereas the conductivity of the β-isomer increased. This suggests that the α-isomer has configuration III, and the β-isomer IV. Thus we now have physico-chemical evidence that the 1:2-hydroxyl groups are in the *cis*-position

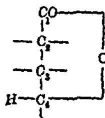


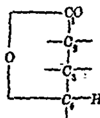


in the α-isomer, *i.e.*, there is now some experimental evidence in support of Hudson's empirical rule. These configurations have been confirmed by further work, *e.g.*, Rüber (1931) found that, in general, *trans*-compounds have a higher molecular refraction than the corresponding *cis*-; the molecular refraction of β-D-glucose is greater than that of the α-isomer, and so agrees with the results obtained by the conductivity experiments. The strongest bit of evidence

for the configurations of the α- and β-isomers has been obtained from X-ray studies of α-D-glucose (see §7f).

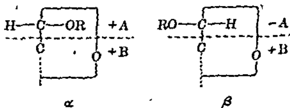
§5. **Hudson's lactone rule.** Hudson (1910) studied the rotation of the lactones derived from the aldonic acids. Using the usual projection formulae, the lactone ring will be on the right or left according as the hydroxyl group on C_4 (*i.e.*, the γ-hydroxyl group) is on the right or left, *i.e.*, according as C_4 has a *dextro* or *lævo* configuration:





From an examination of 24 lactones derived from aldonic acids, and assuming that they were γ-lactones, Hudson concluded that if the lactone ring was on the right, the compound was dextrorotatory; if the ring was on the left, then lævorotatory.

§6. **Hudson's isorotation rules.** Hudson (1909, 1930) applied the rule of optical superposition (§12. I) to carbohydrate chemistry, and his first application was to the problem of the configuration of C_1 in the anomers of aldoses. Hudson pointed out that the only structural difference between the α- and β-anomers (of sugars and glycosides) is the configuration of C_1. Thus, representing the rotation of this terminal group as A and that of the rest of the molecule as B, and then taking the α-anomer as the one with the higher positive rotation (in the D-series) we have:



Molecular rotation of the α-anomer = + A + B
" " " " β- " = − A + B

Thus in every pair of α- and β-anomers the following rules will hold:

Rule 1. The sum of the molecular rotations (2B) will be a constant value characteristic of a particular sugar and independent of the nature of R.

Rule 2. The difference of the molecular rotations (2A) will be a constant value characteristic of R.

As we have seen, the rule of optical superposition does not hold exactly (due to neighbouring action, etc.; see §12. I). In the sugars, however, the rotation of C_1 is affected only to a small extent by changes in the rest of the molecule, and *vice versa*. This is illustrated in the following table, from which it can be seen that the sum of the molecular rotations (2B) for various pairs of glucopyranoside anomers is fairly constant.

C_1 substituent	M_α	M_β	$M_\alpha + M_\beta = 2B$
OH	+202	+34	+236
OCH_3	+309	−66	+243
OC_2H_5	+314	−65·5	+248·5

These isorotation rules have been used to ascertain which of an anomeric pair of glycosides is α and which is β, and to determine the type of glycosidic link in disaccharides and polysaccharides.

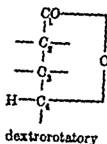
§7. Methods for determining the size of sugar rings. As pointed out previously, Fischer followed Tollens in proposing the γ-oxide ring. There was, however, no experimental evidence for this; the γ-hydroxyl group was chosen as being involved in ring formation by analogy with the ready formation of γ-lactones from γ-hydroxy-acids. The problem was further complicated by the fact that Hudson *et al.* (1915) isolated four galactose penta-acetates, none of which had a free aldehyde group. Furthermore, these four compounds were related to each other as pairs, *i.e.*, there were two α- and two β-isomers. The only reasonable explanation for this was that there are *two* ring systems present, but once again there is no evidence to decide the actual sizes of the rings.

The original experimental approach to the problem of determining the size of the ring present in sugars consisted essentially in studying the methylated sugars. A more recent method uses the methyl glycosides (for this method, see §7g). Since methylation is so important in the original method, the following account describes briefly the methods used.

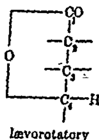
(i) *Purdie's method* (1903). The sugar is first converted corresponding methyl glycoside (methanol and hydrochl

for the configurations of the α- and β-isomers has been obtained from X-ray studies of α-D-glucose (see §7f).

§5. Hudson's lactone rule. Hudson (1910) studied the rotation of the lactones derived from the aldonic acids. *Using the usual projection formulae, the lactone ring will be on the right or left according as the hydroxyl group on C_4 (i.e., the γ-hydroxyl group) is on the right or left, i.e., according as C_4 has a dextro or lævo configuration.*



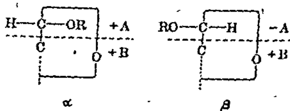
dextrorotatory



lævorotatory

From an examination of 24 lactones derived from aldonic acids, and assuming that they were γ-lactones, Hudson concluded that *if the lactone ring was on the right, the compound was dextrorotatory; if the ring was on the left, then lævorotatory.*

§6. Hudson's isorotation rules. Hudson (1909, 1930) applied the rule of optical superposition (§12. I) to carbohydrate chemistry, and his first application was to the problem of the configuration of C_1 in the anomers of aldoses. Hudson pointed out that *the only structural difference between the α- and β-anomers (of sugars and glycosides) is the configuration of C_1. Thus, representing the rotation of this terminal group as A and that of the rest of the molecule as B, and then taking the α-anomer as the one with the higher positive rotation (in the D-series)* we have:



Molecular rotation of the α-anomer = +A + B
" " " " β- " = −A + B

Thus in every pair of α- and β-anomers the following rules will hold:

Rule 1. The sum of the molecular rotations (2B) will be a constant value characteristic of a particular sugar and independent of the nature of R.

Rule 2. The difference of the molecular rotations (2A) will be a constant value characteristic of R.

As we have seen, the rule of optical superposition does not hold exactly (due to neighbouring action, etc.; see §12. I). In the sugars, however, the rotation of C_1 is affected only to a small extent by changes in the rest of the molecule, and *vice versa*. This is illustrated in the following table, from which it can be seen that the sum of the molecular rotations (2B) for various pairs of glucopyranoside anomers is fairly constant.

C_1 substituent	M_α	M_β	$M_\alpha + M_\beta = 2B$
OH	+ 202	+ 34	+ 236
OCH_3	+ 309	− 66	+ 243
OC_2H_5	+ 314	− 69·5	+ 245·5

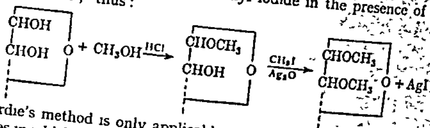
These isorotation rules have been used to ascertain which of an anomeric pair of glycosides is α and which is β, and to determine the type of glycosidic link in disaccharides and polysaccharides.

§7. Methods for determining the size of sugar rings. As pointed out previously, Fischer followed Tollens in proposing the γ-oxide ring. There was, however, no experimental evidence for this; the γ-hydroxyl group was chosen as being involved in ring formation by analogy with the ready formation of γ-lactones from γ-hydroxy-acids. The problem was further complicated by the fact that Hudson *et al.* (1915) isolated four galactose penta-acetates, none of which had a free aldehyde group. Furthermore, these four compounds were related to each other as pairs, *i.e.*, there were two α- and two β-isomers. The only reasonable explanation for this was that there are *two* ring systems present, but once again there is no evidence to decide the actual sizes of the rings.

The original experimental approach to the problem of determining the size of the ring present in sugars consisted essentially in studying the methylated sugars. A more recent method uses the methyl glycosides (for this method, see §7g). Since methylation is so important in the original method, the following account describes briefly the methods used.

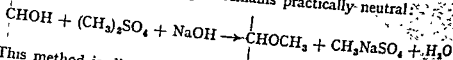
(i) *Purdie's method* (1903). The sugar is first converted into the corresponding methyl glycoside (methanol and hydrochloric acid),

and this is then heated with methyl iodide in the presence of dry silver oxide; thus:



Purdie's method is only applicable to glycosides and other derivatives in which the *reducing group* is missing or has been protected by substitution. Methylation of a free reducing sugar by this method would result in the oxidation of that sugar by the silver oxide.
In certain cases, thallous hydroxide may be used instead of silver oxide (Fear *et al.*, 1926).

(ii) *Haworth's method* (1915). In this method methyl sulphate and aqueous sodium hydroxide are added to a well-stirred sugar solution at such a rate that the liquid remains practically neutral:

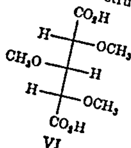


$$\text{CHOH} + (CH_3)_2SO_4 + NaOH \longrightarrow \text{CHOCH}_3 + CH_3NaSO_4 + H_2O$$

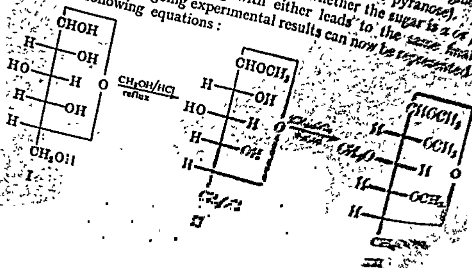
This method is directly applicable to all reducing sugars.
(iii) More recent methods of methylation use sodium and methyl iodide in liquid ammonia, or diazomethane in the presence of moisture.
Having obtained the fully methylated methyl glycoside, the latter is ... ted methyl glycoside, the ... hydrochloric acid, whereby ... methylated sugar ... A study of the oxidation ... methylated sugar then leads to the size of the ring. It should be noted that throughout the whole method, the assumption is made that no methyl groups migrate or that any change in the position of the oxide ring occurs (see, however, later). The number of methyl groups present in the methylated sugar and the various oxidation products are determined by the Zeisel method (see Vol. I). Also, these methyl derivatives are often purified by distillation *in vacuo*.

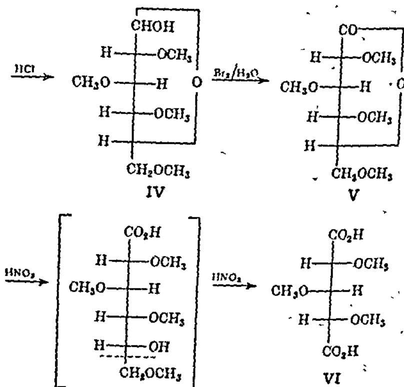
§7a. Pyranose structure. This structure is also sometimes referred to as the δ-oxide or amylene oxide ring. As an example of the method used, we shall consider the case of D(+)-glucose (Haworth and Hirst, 1927). D(+)-Glucose, I, was refluxed in methanol solution in the presence of a small amount of hydrochloric acid, and the methyl D-glucoside, II, so produced was methylated with methyl sulphate in the presence of sodium hydroxide to give methyl tetramethyl-D-glucoside, III, and this, on hydrolysis with dilute hydrochloric acid, gave tetramethyl-D-glucose, IV.

When this was dissolved in water and then oxidised by heating with excess of bromine at 90°, a lactone, V, was isolated, and this, on further oxidation with nitric acid, gave xylotrimethoxyglutaric acid, VI. The structure of this compound is known, since it can be obtained directly by the oxidation of methylated xylose; thus its structure is VI (see also §7d). The structure of this compound is

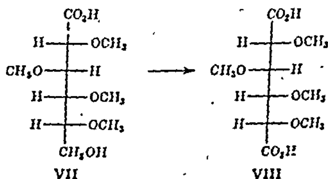


the key to the determination of the size of the ring in the sugar. One of the carboxyl groups in VI must be that which is combined in the formation of the lactone ring in the tetramethylgluconolactone, V. The other carboxyl group is almost certainly the one that has been derived from the CHOH group that is involved in the ring formation in the sugar. Therefore there must be *three* methoxyl groups in the lactone ring. Thus the lactone *cannot* be a γ-lactone, and consequently C_5 must be involved in the ring formation. It therefore follows that the lactone, V, must be 2:3:4:6-tetra-*O*-methyl-D-gluconolactone. Working *backwards* from this compound, then IV must be 2:3:4:6-tetra-*O*-methyl-D-glucose, III methyl 2:3:4:6-tetra-*O*-methyl-D-glucoside, II methyl D-glucopyranoside, and I D-glucopyranose (see §7f for the significance of the term pyranose). It should be noted that the question as to whether the sugar is α or β has been ignored; starting with either leads to the same final results. The foregoing experimental results can now be represented by the following equations:



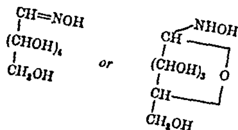


There is a slight possibility that the ring might have been an ε-ring, *i.e.*, the oxide ring involves C_1 and C_4, and that C_5 is converted to the carboxyl group with loss of C_6. Haworth, however, made certain that this was not the case by the following method. Had the ring been 1:6-, then 2:3.4:5-tetramethylgluconic acid, VII, would have been obtained (instead of V). VII was obtained by Haworth *et al.* (1927) from melibiose and gentiobiose (see §§18, 19) and, on oxidation, gave tetramethylsaccharic acid, VIII, and not the dicarboxylic acid, VI.



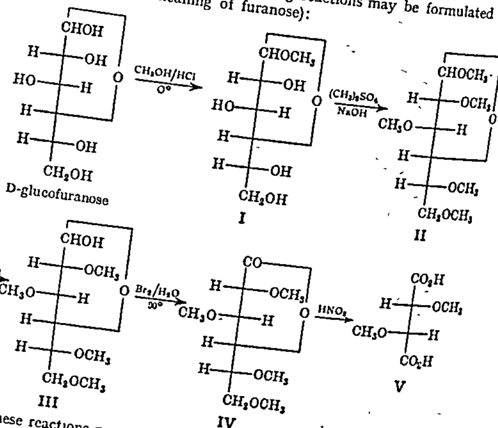
Thus there is a 1:5-ring in the tetramethylgluconolactone, tetra-*O*-methylglucose, methyl tetra-*O*-methylglucoside, methyl glucoside, *and therefore in glucose itself*. This conclusion is based on the assumption that no change in the ring position occurs during the methylation of glucose. Thus glucose is a δ- or pyranose sugar.

By similar methods it has been shown that hexoses and pentoses all possess a pyranose structure. There is also a large amount of evidence to show that the oximes, phenylhydrazones, and osazones of hexoses and pentoses are also cyclic, but there is some evidence, however, for the open-chain formula as well, *e.g.*, the oxime of glucose:



§7b. **Furanose structure.** This structure is also sometimes referred to as the γ-oxide or butylene oxide ring. Fischer (1914) prepared methyl D(+)-glucoside by a slightly modified method, *viz.*, by dissolving D(+)-glucose in methanol, adding one per cent. hydrochloric acid, and then allowing the mixture to stand at 0° (instead of refluxing, as in his first procedure). On working up the product, he obtained a syrup (a crystalline compound was obtained by the first procedure). Fischer called this compound methyl γ-glucoside, and believed it was another isomer of the α and β forms; this is the significance of the symbol γ as used by Fischer. This syrup, however, was subsequently shown to be a mixture of methyl α- and β-glucofuranosides, *i.e.*, this glucoside contained a γ- or 1.4-ring (Haworth *et al.*, 1927). This syrup, I, when completely methylated (methyl sulphate method), gave a methyl tetra-*O*-methyl-D-glucoside, II, and this, on hydrolysis with dilute hydrochloric acid, gave tetra-*O*-methyl-D-glucose, III. On oxidation with bromine water at 90°, III gave a crystalline lactone, IV, and this, when oxidised with nitric acid, gave dimethyl-D-tartaric (dimethoxysuccinic) acid, V. This compound (V) is the only compound of known structure, and is therefore the key to the determination of the size of the ring in the sugar. Working *backwards* from V, then IV is 2:3:5:6-tetra-*O*-methyl-D-gluconolactone, III is 2:3:5:6-tetra-*O*-methyl-D-glucose, II is methyl 2:3:5:6-tetra-*O*-methyl-D-glucoside, and I is methyl D-glucofuranoside. If we write D-glucose as

D-glucofuranose, then the foregoing reactions may be formulated (see §7f for the meaning of furanose):



These reactions prove that I, II, III and IV all contain a γ-oxide ring, *i.e.*, the methyl glucoside, I, *prepared at* 0°, has a 1:4-ring. This then raises the question: What is the size of the ring in glucose itself? Is it 1:4 or 1:5? Preparation of the methyl glucoside at reflux temperature gives the 1:5-compounds (see §7a); preparation at 0° gives the 1:4-compounds. It is therefore not possible to say from these experiments [illegible] in the pyranos[illegible] these two f[illegible] sary to su[illegible] these questions. As we shall see later, the normal form of a sugar is the pyranose structure (see §7f); pyranosides are often referred to as the "normal" glycosides.

By similar methods it has been shown that hexoses and pentoses give methyl glycosides possessing a furanose structure when prepared at 0° (or at room temperature).

§7c. Determination of ring size by means of lactone formation. As we have seen, glycoside formation at *reflux* tempera-

ture leads ultimately to a methylated δ-lactone, whereas at 0° a methylated γ-lactone is obtained Haworth (1927) examined the rates of hydration of these two types of lactones to the open-chain acids; the rates were measured by changes in the rotation or conductivity. Haworth found that the rate of hydration was much faster in one series than in the other, the δ-lactones were converted almost completely to the acids, whereas the γ-lactones were converted at a much slower rate (see Fig. 1) Thus, by comparing

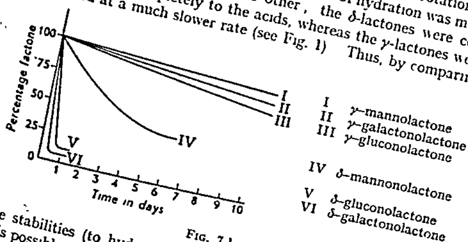
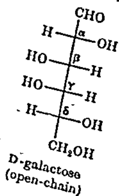
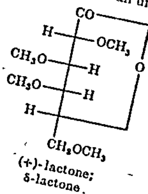


Fig. 7.1.

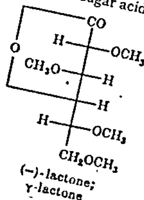
the stabilities (to hydration) of the various methylated lactones, it is possible to say whether the lactone under investigation is γ- or δ-. It is very important to note that this method easily distinguishes a γ- from a δ-lactone, but it does not *prove* one to be γ- and the other δ-. The actual nature of the lactone was proved *chemically*; the fast-changing lactone was shown to be the δ-lactone, and the slow-changing one the γ- (the chemical evidence was obtained by the degradative oxidation already described). However, having once established the relationship between the rate of hydration and the nature of the lactone, *e g.*, in the case of glucose, mannose, galactose and arabinose, the property can *then* be used to determine the size of the ring in an unknown lactone of a sugar acid.



D-galactose (open-chain)



(+)-lactone; δ-lactone.



(−)-lactone; γ-lactone

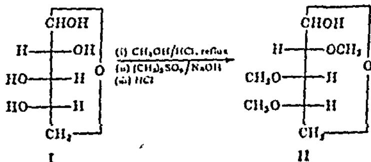
Correlation between the above scheme and Hudson's lactone rule has been demonstrated in certain cases, *e.g.*, galactose. Preparation of the methyl galactoside at reflux temperature, then methylation, hydrolysis, and finally oxidation with bromine water, leads to the formation of a methylated lactone which is dextrorotatory, and since it is a rapidly hydrated lactone, it must be δ-. Preparation of the methyl galactoside at 0°, etc., leads to the formation of a methylated lactone which is lævorotatory and is very stable to hydration. Thus, this lactone will have the ring to the left (Hudson's lactone rule), and hence must be a γ-lactone; at the same time, since it is a slowly hydrated lactone, it must be γ- (see the above formulae).

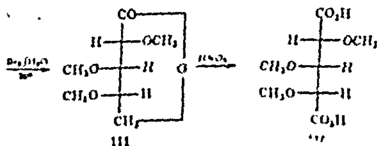
§7d. **Pyranose and furanose structures of pentoses.** The methods used for determining the size of sugar rings have been described with glucose (an aldohexose) as the example. *It is also* instructive to apply these methods to the aldopentoses. L(+)-Arabinose has been chosen as the example, and the following equations and footnotes should now be readily followed:

(i) *Glycoside formation at reflux temperature* (Haworth *et al.*, 1927).

I is L(+)-arabinopyranose, and since it is *dextrorotatory*, the ring has been drawn to the *right*. This way of drawing the projection formula is based on the observation of Haworth and Drew (1926), who pointed out that if a ring in a sugar is 1:5- (*i.e.*, δ-), then Hudson's lactone rule holds good for sugars as for γ-lactones.

II is 2:3:4-tri-*O*-methyl-L-arabinose.





III is 2:3:4-tri-*O*-methyl-L-arabinolactone; it is a δ-lactone as shown by oxidation to IV, and also by the fact that it is of the type that is readily hydrated.

IV is 2:3:4-L-arabinotrimethoxyglutaric acid (this is the key compound).

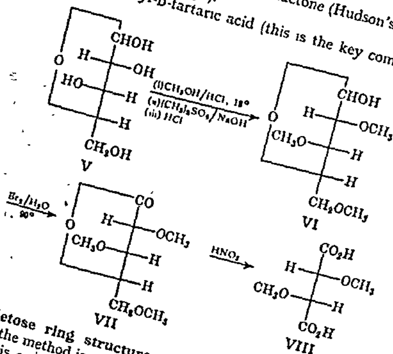
(ii) *Glycoside formation at room temperature* (Haworth *et al.*, 1925, 1927).

V is L-arabinofuranose.

VI is 2:3:5-tri-*O*-methyl-L-arabinose.

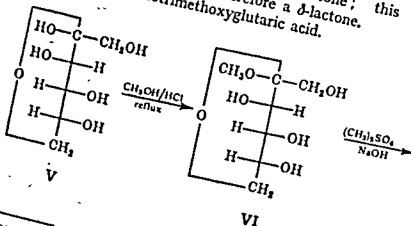
VII is 2:3:5-tri-*O*-methyl-L-arabinolactone (Hudson's lactone rule, and is slow-changing type).

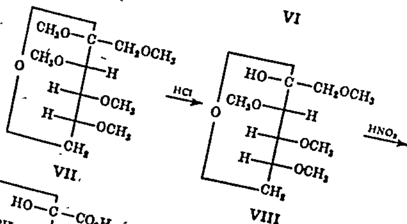
VIII is dimethyl-D-tartaric acid (this is the key compound).

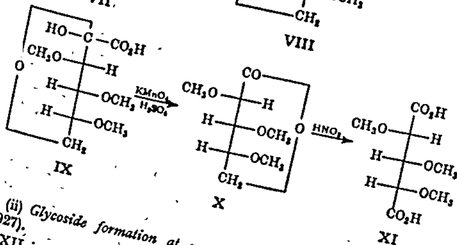


§7e. Ketose ring structures. Only D-fructose will be considered; the method is essentially the same as that for the aldoses, but there is one important variation, and that is in the oxidation of the tetramethylfructose. This cannot be oxidised by bromine water as can the tetramethylaldose; the fructose derivative is first oxidised with dilute nitric acid and then with acid permanganate, and by this means the lactone is obtained. The lactone is then further oxidised by moderately concentrated nitric acid. The following equations and footnotes explain the method, but before giving these, let us first consider the way of writing the projection formula of the ring structure of fructose.

X is 2:3:4-tri-*O*-methyl-D-arabinolactone; this is a quick-changing lactone, and is therefore a δ-lactone.
XI is D-arabinotrimethoxyglutaric acid.

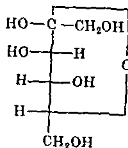


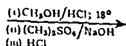


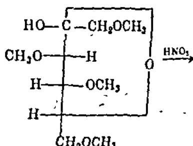


(ii) *Glycoside formation at room temperature* (Haworth *et al.*, 1927).

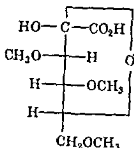
XII is β-D-fructofuranose.
XIII is 1:3:4:6-tetra-*O*-methyl-β-D-fructose.
XIV is 3:4:6-tri-*O*-methyl-β-D-fructuronic acid (as lactol).

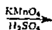


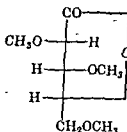


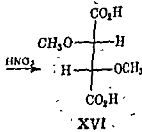


XII XIII





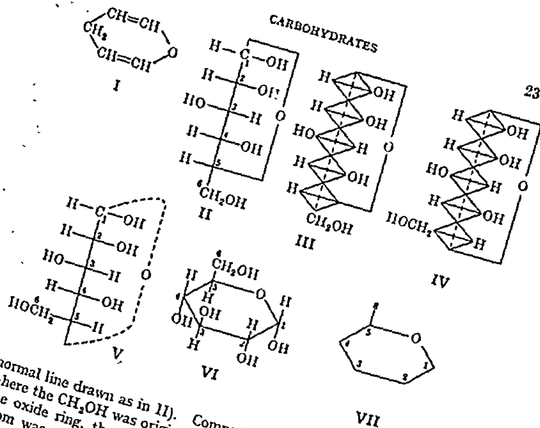




XIV XV

XV is 2·3:5-tri-*O*-methyl-D-arabinolactone; this is a slow-changing lactone, and so is γ-.
XVI is dimethyl-L-tartaric acid.

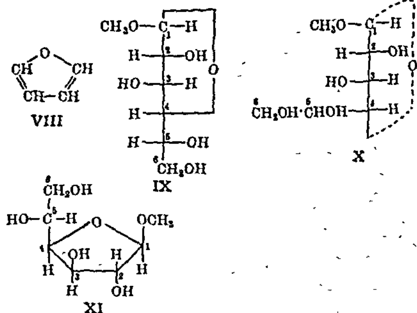
§7f. Conclusion. From the foregoing account it can be seen that the sugars exist as ring structures and not as open chains. Haworth (1926) therefore proposed a hexagonal formula for δ-sugars based on the pyran ring, I. The problem now is to convert the conventional plane-diagrams that we have been using into the *pyranose* formula. Let us take α-D-glucopyranose, II, as our example. The conventional tetrahedral diagram of II is III (see §5. II). Examination of III shows that the point of attachment of the oxide ring at C_1 is *below* the plane of the paper, and that at [illegible] and the oxide ring will now be *perpendicular* to the plane of the paper, *i.e.*, perpendicular to the plane containing all the other groups (these all lie in a plane above the plane of the paper). The conventional plane-diagram of IV is V, but in order to emphasise the fact that the oxide ring is actually perpendicular to the plane of the paper, the part of the ring lying below the plane of the paper is shown by a broken line (the true plane-diagram should have a



normal line drawn as in *II*). Comparison of V with II shows that where the CH_2OH was originally is now the point of attachment of the oxide ring, the CH_2OH occupying the position where the H atom was, and the latter now where the oxide ring was. Thus, if we consider the conversion of II into V without first drawing III and IV, then in effect *two* Walden inversions have been effected, and consequently the original configuration is retained. V is now transformed into the perspective formula VI by twisting V so that the oxide ring is perpendicular to the plane of the paper and all the other groups are joined to bonds which are parallel to the plane of the paper. By convention, C_1 is placed to the right and the oxygen atom at the right-hand side of the part of the ring furthest from the observer. Sometimes the lower part of the ring, which represents the part nearest to the observer, is drawn in thick lines. Thus, to change V into VI, first draw the hexagon as shown in VI, and then place all the groups on the left-hand side in V above the plane of the ring in VI; all those on the right-hand side in V are placed below the plane of the ring in VI. VII represents a "short-hand representation" of D-glucose.

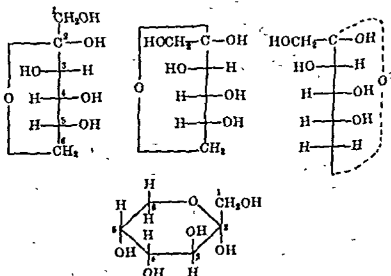
In a similar manner, Haworth proposed a five-membered ring for γ-sugars based on the furan ring, VIII. Using the above scheme of transformation, the plane-diagram of methyl β-D(+)-gluco-

furanoside, IX, is first changed into X (*two* changes are carried out), and then X is twisted so as to be represented by XI, in which the oxygen atom is furthest from the observer.

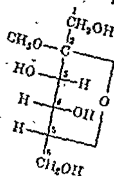


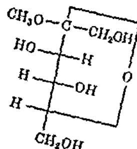
Two other examples which illustrate the conversion into the perspective formula are:

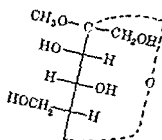
(i) **α-D(−)-fructopyranose.**

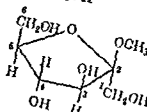


(ii) Methyl β-D(+)-fructofuranoside.





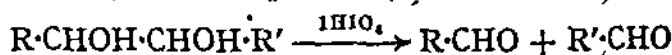




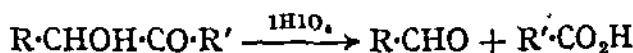
Actual size of sugar rings. Since glycoside formation under different conditions gives compounds containing different sized rings, the important question then is: What is the size of the ring in the original sugar? Oxidation of an aldose with hypobromite produces an unstable δ-lactone; this is the first product, but slowly changes into the stable γ-lactone (Hudson, 1932). It therefore follows that the size of the ring in *normal* sugars is pyranose. By analogy, ketoses are also believed to exist normally as pyranose compounds. This pyranose structure has been confirmed by X-ray analysis of various crystalline monosaccharides (Cox, 1935). McDonald *et al.* (1950) examined α-D-glucose by X-ray analysis, and confirmed the presence of the six-membered ring, the configuration as found chemically, and also the *cis* arrangement of the 1:2-hydroxyl groups in the α-form. Eiland *et al.* (1950) subjected difructose strontium chloride dihydrate to X-ray analysis, and showed the presence of a six-membered ring, and confirmed the configuration found chemically. It might be noted here that furanose sugars have not yet been isolated, but some furanosides have. It is also interesting to note that apparently fructose and ribose *always* occur in compounds as the furanose structure.

§7g. More recent methods for determining the size of the ring in sugars. These methods make use of the fact that periodic

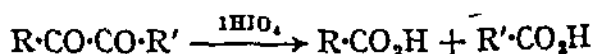
acid splits 1:2-glycols (Malaprade, 1928); thus periodic acid splits the following types of compounds (see also Vol. I):



$$R{\cdot}CHOH{\cdot}CHOH{\cdot}R' \xrightarrow{1HIO_4} R{\cdot}CHO + R'{\cdot}CHO$$

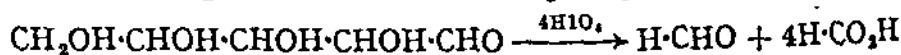


$$R{\cdot}CHOH{\cdot}CO{\cdot}R' \xrightarrow{1HIO_4} R{\cdot}CHO + R'{\cdot}CO_2H$$



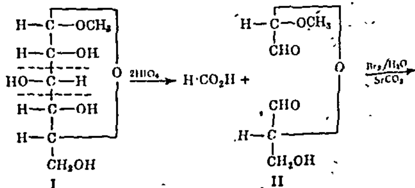
$$R{\cdot}CO{\cdot}CO{\cdot}R' \xrightarrow{1HIO_4} R{\cdot}CO_2H + R'{\cdot}CO_2H$$

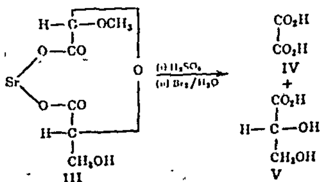
Thus a *free* sugar is broken down completely, *e.g.*,



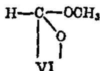
$$CH_2OH{\cdot}CHOH{\cdot}CHOH{\cdot}CHOH{\cdot}CHO \xrightarrow{4HIO_4} H{\cdot}CHO + 4H{\cdot}CO_2H$$

In all of these reactions, one molecule of periodic acid is used for each pair of adjacent alcoholic groups (or oxo groups). Thus, by estimating the periodic acid used, and the formic acid and formaldehyde formed, the number of *free* adjacent hydroxyl groups in a sugar can be ascertained. Hudson (1937, 1939) oxidised "normal" methyl α-D-glucoside, I, with periodic acid, and found that two molecules of periodic acid were consumed, and that one molecule of formic acid was produced. It should be noted that although periodic acid can completely degrade a *free* sugar, the oxide ring in glycosides is sufficiently stable to resist opening by this reagent. The first pro-



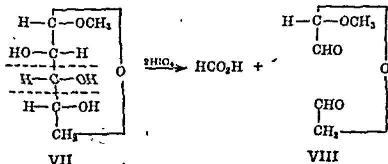


duct of oxidation of methyl α-D-glucoside was D′-methoxy-D-hydroxymethyldiglycolaldehyde, II, and this, on oxidation with bromine water in the presence of strontium carbonate, gave the crystalline salt, III. III, on acidification with sulphuric acid (for hydrolysis), followed by further oxidation with bromine water, gave oxalic acid, IV, and D(−)-glyceric acid, V. Isolation of II, III, IV and V indicates that the ring in I is δ-; this is also supported by the fact that only one carbon atom was eliminated as formic acid, and that two molecules of periodic acid were consumed. By similar experiments, it has been shown that all methyl α-D-hexosides of the "normal" type consume *two* molecules of periodic acid and produce *one* molecule of formic acid, and all also give products II, III, IV and V. Thus all these hexosides must be six-membered rings, and also it follows that all "normal" methyl α-pyranosides have the same configuration for C_1; this has already been shown to be VI.



Similarly, all β-compounds, on oxidation with periodic acid, give the stereoisomer of II, *i.e.*, L′-methoxy-D-hydroxymethyldiglycolaldehyde.

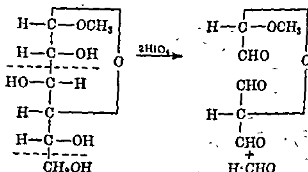
Aldopentopyranosides also give similar products as those obtained from the aldohexopyranosides, *e.g.*, methyl α-D-arabinopyranoside, VII, gives D′-methoxydiglycolaldehyde, VIII. Since



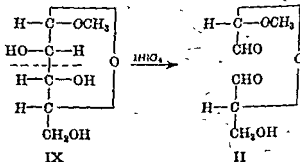
all methyl α-D-aldopentopyranosides give the *same* diglycolaldehyde, they too have the same configuration for C_1, *viz.*, VI.

When hexofuranosides, *i.e.*, the "abnormal" glycosides, are oxidised with periodic acid, *two* molecules of acid are consumed and one molecule of *formaldehyde* is formed. These results are

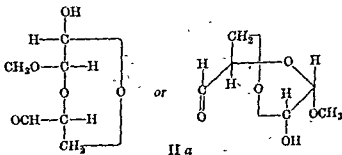
in keeping with the presence of a five-membered ring, *e.g.*, methyl α-D-glucofuranoside.



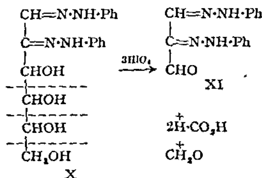
Oxidation of methyl α-D-arabinofuranoside, IX, consumes *one* molecule of periodic acid, and no carbon atom is eliminated (either as formaldehyde or formic acid); thus the ring is five-membered. Furthermore, since the dialdehyde II obtained is the same as that from methyl α-D-glucopyranoside, I, the configuration of C_1 is the same in both I and IX.



There appears to be some doubt about the structure of II. Various formulae have been proposed (Hurd *et al.*, 1953; Smith *et al.*, 1955), and Mester *et al.* (1957) have obtained evidence that of these structures the cyclic hemiacetal (*IIa*) is the most likely.



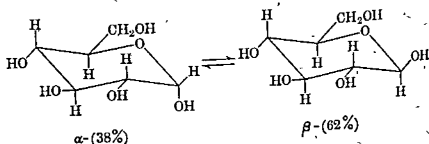
Hough *et al.* (1956) have carried out periodate oxidations on phenylosazones of reducing monosaccharides (X) and obtained formaldehyde, formic acid and mesoxalaldehyde 1:2-bisphenylhydrazone (XI). These authors found that XI is obtained from



all monosaccharides in which C_3 and C_4 are free, and 1 molecule of formaldehyde from the terminal CH_2OH group when this is free. They also showed that the osazones of the disaccharides maltose (§15), cellobiose (§16), and lactose (§17) did not give XI but did give formaldehyde. Thus C_3 or C_4 are linked in these disaccharides. On the other hand, the oxidation of the osazone of melibiose (§18) gave XI but no formaldehyde; thus C_6 is linked in this molecule. These oxidations therefore offer a means of differentiating between the two types of disaccharides.

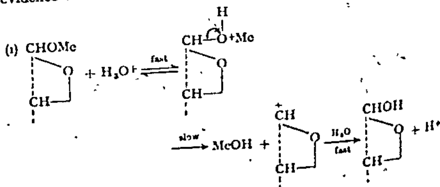
other by ring conversion; *cf. cyclo*hexane, §11. IV). Reeves has found that almost all of the D-pyranoses have the C1 conformation and L-pyranoses the 1C. D-Idose is an exception in that it has the 1C conformation, and D-altrose (and D-lyxose) is an equilibrium mixture of C1 and 1C.

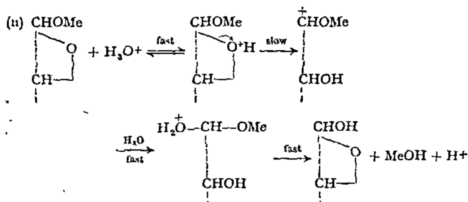
As we have seen (§2), D-glucopyranose is an equilibrium mixture (in solution) of the α- and β- anomers: the conformations of these are:



We have also seen that the more stable isomer is the one with the larger number of equatorial substituents, and so the β-form can be expected to be more stable than the α-. Whiffen *et al.* (1954) have used infrared spectroscopy to distinguish between α- and β-anomers; the absorption maxima depend on the axial or equatorial conformation of hydroxyl groups.

In general, β-anomers are more reactive than α-, *e.g.*, Bunton *et al.* (1954) have shown that acid-catalysed hydrolysis proceeds more rapidly for β-methyl pyranosides than for the corresponding α-compounds. According to these authors (1955), the hydrolysis proceeds by a unimolecular decomposition of the conjugate acids of the pyranosides. The rate-determining step, however, may be formulated in two ways, both of which are consistent with the evidence available at present.





On the other hand, axial hydroxyl groups are less reactive (to esterification and hydrolysis reactions) than equatorial groups (§12. IV). In β-pyranosides, the methoxyl group is equatorial and so mechanism (i) would be more in keeping with the fact that β-anomers are more readily hydrolysed than α- (in which the methoxyl group is axial).

§8. *iso*Propylidene derivatives of the monosaccharides. Sugars condense with anhydrous acetone in the presence of hydrogen chloride, sulphuric acid, etc., at room temperature to form mono- and di-*iso*propylidene (or acetone) derivatives. These are stable towards alkalis, but are readily hydrolysed by acids. In the di-*iso*propylidene derivatives, one *iso*propylidene group is generally removed by hydrolysis more readily than the other, and thus by controlled hydrolysis it is possible to isolate the mono-*iso*propylidene derivative, *e.g.*, di-*iso*propylideneglucose may be hydrolysed by acetic acid to the mono-derivative.

The structures of these *iso*propylidene derivatives have been determined by the methods used for the sugars themselves, *i.e.*, the compound is first methylated, then hydrolysed to remove the acetone groups, and the product finally oxidised in order to ascertain the positions of the methyl groups. Let us consider D-glucose as an example. This forms a di-*iso*propylidene derivative, I, which is non-reducing; therefore C_1 is involved in the formation of I. On methylation, I forms a monomethyldi-*iso*propylidene glucose, II, and this, on hydrolysis with hydrochloric acid, gives a monomethylglucose, III. Hydrolysis of I with acetic acid produces a mono-*iso*propylideneglucose, IV, which is also non-reducing. Thus C_1 in IV must be combined with the *iso*propylidene radical. Methylation of IV, followed by hydrolysis, gives a trimethylglucose, V. Methylation of V gives a methyl tetramethylglucoside, and this,

on hydrolysis, gives 2:3:5:6-tetra-*O*-methyl-D-glucose, VI, a *known* compound (see §7b). Thus V must be 2:3:5-, 2:3:6-, or 3:5:6-tri-*O*-methyl-D-glucose. Now V forms an osazone without loss of any methyl group; therefore C_2 cannot have a methoxyl group attached to it, and so V must be 3:5:6-tri-*O*-methyl-D-glucose. Thus one *iso*propylidene radical in di-*iso*propylideneglucose, I, must be 3:5-, 3:6-, or 5:6-. Monomethylglucose, III, on methylation followed by hydrolysis, gives 2:3:4:6-tetra-*O*-methyl-D-glucose, VII, a *known* compound (see §7a). Hence III must be 2-, 3-, 4-, or 6-*O*-methyl-D-glucose. Since III gives sodium cyanate when subjected to the Weerman test (see §11), it therefore follows that C_2 has a free hydroxyl group. Oxidation of III with nitric acid produces a monomethylsaccharic acid; therefore C_6 cannot have a methoxyl group attached to it. This monomethylsaccharic acid forms a lactone which behaves as a γ-lactone; therefore a methoxyl group cannot be at C_4. Thus, by the process of elimination, this monomethylglucose, III, must be 3-*O*-methyl-D-glucose. It therefore follows that the two *iso*propylidene groups in the di-*iso*propylidene derivative must be 1:2- and 5:6-, the ring being furanose, and the mono-*iso*propylidene derivative being 1:2-. The foregoing reactions can be written as on opposite page:

α-D(+)-glucose $\xrightarrow[HCl]{(CH_3)_2CO}$ I $\xrightarrow[NaOH]{(CH_3)_2SO_4}$ II

I $\xrightarrow{CH_3\cdot CO_2H}$ IV

II $\xrightarrow{HCl}$ III

IV $\xrightarrow{(i)\ (CH_3)_2SO_4;\ (ii)\ HCl}$ V

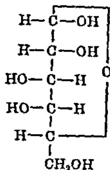
III $\xrightarrow{(i)\ (CH_3)_2SO_4;\ (ii)\ HCl}$ VII

V $\xrightarrow{(i)\ (CH_3)_2SO_4;\ (ii)\ HCl}$ VI

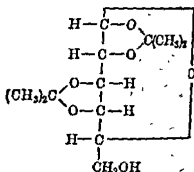
As a result of much experimental work (of the foregoing type), it has been found that acetone usually condenses with *cis*-hydroxyl groups on *adjacent* carbon atoms, the condensation occurring in such a way as to favour the formation of the di-*iso*propylidene

derivative. For this to occur, the ring often changes size, *e.g.*,

pyranose, IX. On the other hand, in α-D-glucopyranose, only



VIII



IX

the two hydroxyl groups on C_1 and C_2 are in the *cis* position, and thus, in order to form the *di-iso*propylidene derivative, the ring

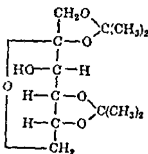
the latter producing 1:2-5:6-

(I). The mono-derivative is

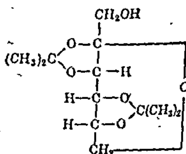
(IV). Fructose can form *two*

both contain the pyranose

ring.



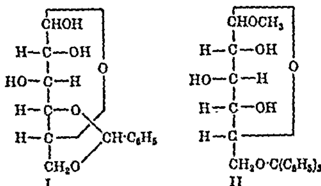
1 2-4:5-



2 3-4:5-

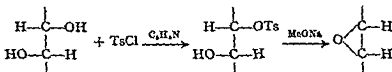
§9. Other condensation products of the sugars. Not only does acetone condense with sugars, but so do other oxo compounds such as formaldehyde, acetaldehyde and benzaldehyde. Benzaldehyde condenses with two *cis* hydroxyl groups on *alternate* carbon atoms, *e.g.*, glucose forms 4:6-*O*-benzylidene-D-glucopyranose, I.

Triphenylmethyl chloride reacts with sugars to form triphenylmethyl ethers; these are usually known as *trityl* derivatives.

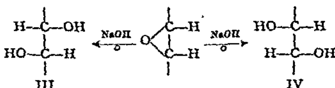


Trityl ethers are formed much faster with primary alcoholic groups than with secondary, *e.g.*, methyl glucopyranoside reacts with triphenylmethyl chloride in pyridine solution to form methyl 6-tritylglucopyranoside, II.

p-Toluenesulphonyl chloride (represented as TsCl in the following equations) reacts with sugars in the presence of pyridine to form *tosyl* esters. These esters usually produce **epoxy-sugars (anhydro sugars)** when hydrolysed with sodium methoxide in the cold, provided that there is a free hydroxyl group on an adjacent carbon atom and that this hydroxyl and the tosyl group are *trans* to each other; *ring formation takes place with a Walden inversion at the tosyl carbon:*



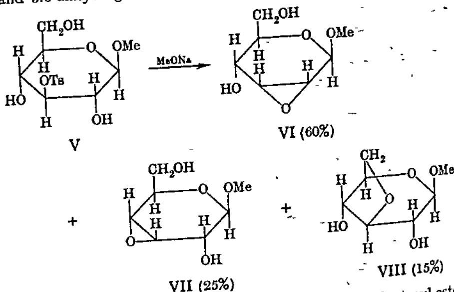
On hydrolysis with alkali, these anhydro sugars form a mixture of *two* sugars, inversion occurring at either carbon when the epoxide ring opens (see §5. IV).



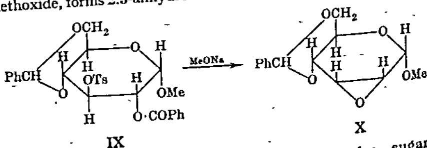
In III the configurations of the two carbon atoms are the same as in the original sugar, but in IV *both* configurations are inverted (to form a new sugar).

When the tosyl group is *trans* to two hydroxyl groups (on adjacent carbon atoms), *two* anhydro sugars are formed. At the same time, however, *larger* epoxide rings may be produced *without* inversion, *e.g.*, Peat *et al.* (1938) treated 3-tosyl methyl *β*-glucoside (V) with

sodium methoxide and obtained a mixture of 2:3-anhydroalloside (VI; with inversion), 3:4-anhydroalloside (VII; with inversion), and 3.6-anhydroglucoside (VIII; no inversion).

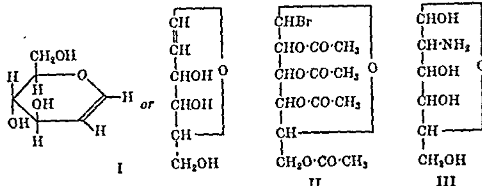


It is possible, however, by using suitable derivatives of a tosyl ester to obtain only one anhydro sugar, *e.g.*, 2-benzoyl-3-tosyl 4.6-benzylidene methyl α-glucoside (IX), on treatment with sodium methoxide, forms 2:3-anhydro 4.6-benzylidene methyl α-alloside (X).



§10. Glycals and glycosamines and anhydro sugars. Glycals are sugar derivatives which have a pyranose ring structure and a double bond between C_1 and C_2, *e.g.*, D-glucal is I. Glycals may be prepared by reducing acetobromo compounds (see §24) with zinc dust and acetic acid, *e.g.*, D-glucal from tetra-*O*-acetyl-D-glucopyranosyl bromide, II, followed by hydrolysis of the acetyl groups.

Glycosamines are amino-sugars in which a hydroxyl group has been replaced by an amino-group. All naturally occurring amino-sugars are hexoses, and the amino-group always occurs on

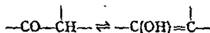


C_2, *e.g.*, glucosamine, which occurs in chitin, is 2-aminoglucose, III (see also §23).

Anhydro sugars. These may be regarded as being derived from monosaccharides by the elimination of a molecule of water to form an epoxide. The size of the oxiran ring varies from 1:2- to 1:6-. The 1:2-anhydro sugars are commonly known as α-glycosans, and may be prepared in various ways, *e.g.*, by heating a sugar under reduced pressure (Pictet *et al.*, 1920). A general method of producing the ethylene oxide series is by the hydrolysis of suitable tosyl esters (see §9).

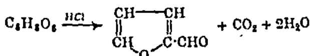
§11. Vitamin C or L-ascorbic acid. Ascorbic acid is very closely related to the monosaccharides, and so is conveniently dealt with here. Hawkins (1593) found that oranges and lemons were effective for treating scurvy, a disease particularly prevalent among seamen. The first significant step in elucidating the nature of the compound, the absence of which from the diet caused scurvy, was that of Holst and Frolich (1907), who produced experimental scurvy in guinea-pigs. Then Szent-Györgi (1928) isolated a crystalline substance from various sources, *e.g.*, cabbages, paprika, etc., and found that it had antiscorbutic properties. This compound was originally called *hexuronic acid*, and later was shown to be identical with vitamin C, m.p. 192°, $[\alpha]_D$ of +24°.

The structure of vitamin C was elucidated by Haworth, Hirst and their co-workers (1932, 1933). The molecular formula was shown to be $C_6H_8O_6$, and since the compound formed a monosodium and monopotassium salt, it was thought that there was a carboxyl group present. Vitamin C behaves as an unsaturated compound and as a strong reducing agent; it also forms a phenylhydrazone and gives a violet colour with ferric chloride. All this suggests that a *keto-enol system is present, i.e.*,

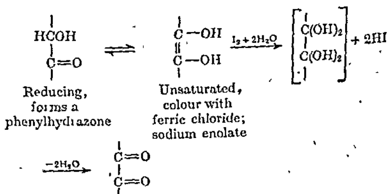


$$-CO-\overset{|}{C}H- \rightleftharpoons -C(OH)=\overset{|}{C}-$$

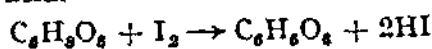
The presence of an aldehyde group was excluded by the fact that vitamin C does not give the Schiff reaction. Now, when boiled with hydrochloric acid, ascorbic acid gives a quantitative yield of furfuraldehyde:



This reaction suggests that ascorbic acid contains at least five carbon atoms in a straight chain, and also that there are a number of hydroxyl groups present (*cf.* the pentoses). Aqueous iodine solution oxidises ascorbic acid to dehydroascorbic acid, two atoms of iodine being used in the process and two molecules of hydrogen iodide are produced; the net result is the removal of two hydrogen atoms from ascorbic acid. Dehydroascorbic acid is neutral, and behaves as the lactone of a monobasic hydroxy-acid; and on reduction with hydrogen sulphid
verted into ascorbic acid. Sinc
may be carried out with "mild"
that since the oxidation product,
then ascorbic acid itself is a lactone and *not* an acid as suggested previously. Since, however, ascorbic acid can form salts, this property must still be accounted for. One reasonable possibility is that the salt-forming property is due to the presence of an *enol* group, the presence of which has already been indicated. Thus all the preceding reactions can be explained by the presence of an α-hydroxyketone grouping in ascorbic acid:

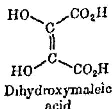


The final result is the removal of two hydrogen atoms to form dehydroascorbic acid.



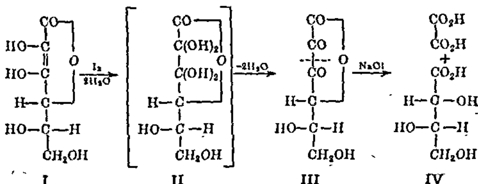
$$C_6H_8O_6 + I_2 \rightarrow C_6H_6O_6 + 2HI$$

Although all these reactions may appear to be speculative, they are known to occur with dihydroxymaleic acid; hence by analogy with this compound, the explanation offered for the reactions of ascorbic acid is very strongly supported.



Dihydroxymaleic acid

When dehydroascorbic acid is oxidised with sodium hypoiodite, oxalic and L-threonic acids are produced in quantitative yields (Hirst, 1933). L-Threonic acid, IV, was identified by methylation and then conversion into the crystalline amide; this compound was shown to be identical with tri-*O*-methyl-L-threonamide (obtained from L-threose). Further evidence for the nature of product IV is given by the fact that on oxidation with nitric acid it gave D(+)-tartaric acid. The formation of oxalic and L-threonic acids suggests that dehydroascorbic acid is III, the lactone of 2:3-diketo-L-gulonic acid. Hence, if we assume that I is the structure of ascorbic acid, the foregoing reactions may be formulated as follows, dehydroascorbic acid being formed *via* II.

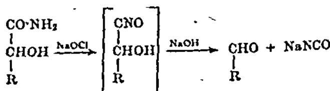


The ring in ascorbic acid has been assumed to be five- and not six-membered, because the lactone (*i.e.*, ascorbic acid) is stable towards alkali (*cf.* §7c). In actual fact, however, the same final products would also have been obtained had the ring been six-membered. It must therefore be admitted that the weakness of the above proof of structure lies in the evidence used for ascertaining the size of the ring. Structure I, however, has been amply confirmed by other analytical evidence. Diazomethane converts ascorbic acid into dimethylascorbic acid (V); these two methoxyl

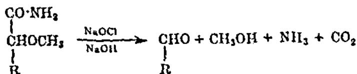
groups are most likely on C_2 and C_3, since diazomethane readily methylates acidic (in this case, enolic) hydroxyl groups. This dimethyl derivative is neutral, and dissolves in aqueous sodium hydroxide to form a sodium salt without the elimination of a methyl group; thus there cannot be a carbomethoxyl group present, and so it is most likely that two enolic hydroxyl groups are present (Hirst, 1933). Furthermore, the formation of the sodium salt from the neutral compound suggests the opening of a lactone ring (the two enolic groups are now methylated and so cannot form a sodium salt). The similarity in structure between ascorbic acid and its dimethyl derivative is shown by the fact that the absorption spectra of both are similar. When this dimethyl derivative is methylated with methyl iodide in the presence of dry silver oxide (Purdie

follows that two alcoholic groups are present in dimethyl-
acid. Ozonolysis of this tetramethyl compound produces *one* neutral substance containing th
its precursor. Since ozonolysi
results in scission of that bond,
in the tetramethyl compound t
(VII). This ozonised product, o...
gives oxalic acid and dimethyl-L-threonic acid (VII). These products contain three carboxyl groups in all, and since ozonolysis of a double bond produces only two, the third carboxyl group must have already been present *as a lactone* in order that ascorbic acid should behave as a neutral compound.

The key to the size of the ring in ascorbic acid is the structure of this dimethyl-L-threonic acid, the nature of which has been ascertained as follows. On methylation, followed by conversion to the amide, dimethyl-L-threonic acid gives trimethyl-L-threonamide. Thus this dimethyl compound, which was unknown when isolated, is a dimethyl-L-threonic acid; but where are the two methoxyl groups? Their positions were ascertained by means of the **Weerman test.** This test is used for showing the presence of a *free* hydroxyl group in the α-position to an amide group, *i.e.*, in an α-hydroxy-amide. Treatment of a methylated hydroxy-amide



with alkaline sodium hypochlorite gives an aldehyde and *sodium cyanate* if there is a *free* hydroxyl group on the α-carbon atom. If there is no free hydroxyl group on the α-carbon atom, *i.e.*, this atom is attached to a methoxyl group, then treatment with alkaline sodium hypochlorite produces an aldehyde, methanol, ammonia and carbon dioxide.



The dimethylthreonic acid obtained from the ozonised product was converted into the amide (IX), and this, when subjected to the Weerman test, gave sodium cyanate as one of the products. Thus this dimethylthreonic acid contains a free α-hydroxyl group, and consequently must be 3:4-di-*O*-methyl-L-threonic acid, VIII. Therefore the lactone ring in ascorbic acid must be γ-, since a δ-lactone could not have given VIII (actually, 2:4-di-*O*-methyl-L-threonic acid would have been obtained). The amide IX was also obtained, together with oxamide, by the action of ammonia in methanol on the ozonised product, VII. All the foregoing facts can be represented by the following equations:

I: CO; HO—C; HO—C; H—C (O ring); HO—C—H; CH_2OH

$\xrightarrow{CH_2N_2}$

V: CO; CH_3O—C; CH_3O—C; H—C (O ring); HO—C—H; CH_2OH

$\xrightarrow[Ag_2O]{CH_3I}$

VI: CO; CH_3O—C; CH_3O—C; H—C (O ring); CH_3O—C—H; CH_2OCH_3

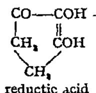
$\xrightarrow{O_3}$

VII: CO; CH_3O—C=O; CH_3O—C=O; H—C (O ring); CH_3O—C—H; CH_2OCH_3

VII $\xrightarrow{NH_3/CH_3OH}$ IX: $CONH_2$–$CONH_2$ + $CONH_2$; H—C—OH; CH_3O—C—H; CH_2OCH_3

VII $\xrightarrow{Ba(OH)_2}$ VIII: CO_2H–CO_2H + CO_2H; H—C—OH; CH_3O—C—H; CH_2OCH_3

An interesting point about ascorbic acid is that it is *not* reduced by lithium aluminium hydride (Petuely *et al.*, 1952). Thus ascorbic acid does not contain a "normal" carbonyl group. It has now been shown that all **reductones** are not reduced by lithium



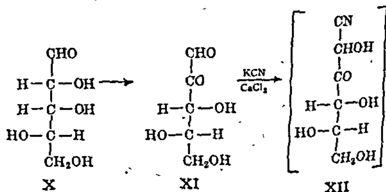
reductic acid

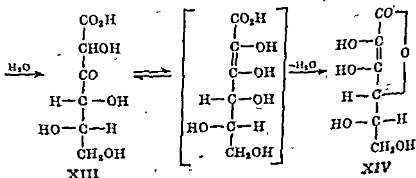
aluminium hydride. Reductions are compounds which contain the ene-α-diol-α-carbonyl grouping, —CO—C(OH)=C(OH)—, and examples of reductiones are ascorbic and reductic acids.

Synthesis of ascorbic acid. Many methods of synthesising

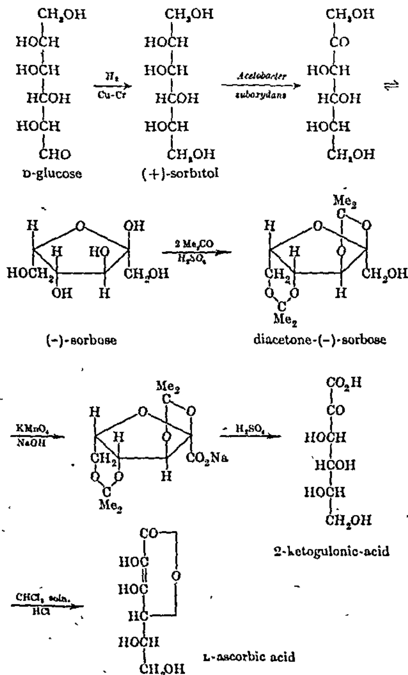
[illegible]

hydrochloric acid), and XI, on treatment in an atmosphere of nitrogen with aqueous potassium cyanide containing calcium chloride, gave the β-keto-cyanide XII, which hydrolyses spon-





taneously into *pseudo*-L-ascorbic acid, XIII. This, on heating for 26 hours with 8 per cent. hydrochloric acid at 45–50°, gave a quantitative yield of L(+)-ascorbic acid, XIV.



Ascorbic acid is now synthesised commercially by several methods, *e.g.*, D-glucose is catalytically hydrogenated to (+)-sorbitol which is then converted into (−)-sorbose by microbiological oxidation (using *Acetobacter suboxydans* or *Acetobacter xylinum*). (−)-Sorbose can be oxidised directly to 2-keto-(−)-gulonic acid with nitric acid, but the yield is less than when the oxidation is carried out as shown above. Nitric acid oxidises other alcohol groups besides the first, but by protecting these by means of 2:3-4:6-di-*iso*propylidene formation (§8), the yield of the gulonic acid is higher. The gulonic acid is then dissolved in mixed solvents (of which chloroform is the main constituent) and hydrogen chloride passed in. The product, L-ascorbic acid, is then finally purified by charcoaling.

DISACCHARIDES

§12. Introduction. The common disaccharides are the dihexoses, and these have the molecular formula $C_{12}H_{22}O_{11}$. Just as methanol forms methyl glycosides with the monosaccharides, so can other hydroxy compounds also form glycosides. The monosaccharides are themselves hydroxy compounds, and so can unite with other monosaccharide molecules to form glycosidic links. Study of the disaccharides (of the dihexose type) has shown that three types of combination occur in the natural compounds:

(i) The two monosaccharide molecules are linked through their reducing groups, *e.g.*, sucrose.
(ii) C_1 of one molecule is linked to C_4 of the other, *e.g.*, maltose.
(iii) C_1 of one molecule is linked to C_6 of the other, *e.g.*, melibiose.

Since the glycosidic link may be α or β, then different stereoisomeric forms become possible for a given pair of hexoses. In group (i), there are [illegible]
and β_1-β_2. [illegible]
molecule is f[illegible]
bilities: α_1-[illegible]
involved in glycosidic formation, the resultant [illegible]
non-reducing. In groups (ii) and (iii), since *one* reducing group is free, the resultant disaccharide will be reducing, and can exist in two forms, the α- and β- [illegible]

General p[illegible]
dilute acids, [illegible]
One of the e[illegible]
was by fractional crystallisation; the separation and identification

is now carried out by means of partition chromatography. When the constituents have been identified, the next problem is to ascertain which hydroxyl group of the molecule acting as the alcohol (*i.e.*, the aglycon; §3) is involved in forming the glycosidic link. This is done by completely methylating the disaccharide; the methyl glycoside (of a reducing sugar) cannot be prepared by means of methanol and hydrochloric acid, since this will lead to hydrolysis of the disaccharide. Purdie's method cannot be used for reducing disaccharides, since these will be oxidised (see §7). The only satisfactory way is Haworth's method, and to ensure complete methylation, this may be *followed* by the Purdie method. The methylated disaccharides are then hydrolysed, and the methylated monosaccharides so obtained are investigated by the oxidation methods described previously (see §§7a, 7b, 7e). Reducing disaccharides are also oxidised to the corresponding bionic acid, this is then fully methylated, hydrolysed, and the methylated monosaccharide molecules examined. By this means the hydroxyl group involved in the glycosidic link and the size of the oxide ring are ascertained.

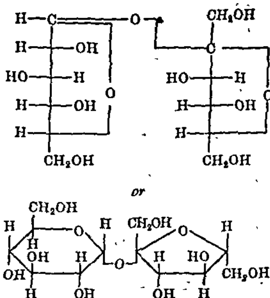
The final problem is to decide whether the glycosidic link is α or β. This is done by means of enzymes, maltase hydrolysing α-glycosides, and emulsin β-glycosides (*cf.* §3). In non-reducing sugars, the problem is far more difficult since the links α_1–α_2, α_1–β_2, β_1–α_2 would *all* be hydrolysed by maltase. Consideration of the optical rotations has given information on the nature of the link (*cf.* §6). Finally, a number of disaccharides have been synthesised, the acetobromo-sugars being the best starting materials (see §24).

§13. Sucrose. Sucrose has been shown to be α-D-glucopyranosyl-β-D-fructofuranoside. Sucrose is hydrolysed by dilute acids or by the enzyme invertase to an equimolecular mixture of D(+)-glucose and D(−)-fructose. Methylation of sucrose (Haworth method) gives octa-*O*-methylsucrose and this, on hydrolysis with dilute hydrochloric acid, gives 2:3:4:6-tetra-*O*-methyl-D-glucose and 1:3:4:6-tetra-*O*-methyl-D-fructose. The structures of these compounds were determined by the oxidation methods previously described (see §§7a, 7e). Thus glucose is present in the pyranose form, and fructose as the furanose.

Since sucrose is a non-reducing sugar, both glucose and fructose must be linked *via* their respective reducing groups. The stereochemical nature of the glycosidic link may be any one of the four possibilities discussed (see §12), but the evidence indicates that it is α-glucose linked to β-fructose. Maltase hydrolyses sucrose;

therefore an α-link is present. Furthermore, since the mutarotation of the glucose produced is in a downward direction, it therefore follows that α-glucose is liberated at first. The mutarotation of fructose is too rapid to be followed experimentally, and hence the nature of the link in this component remains to be determined.

hydrolyses methyl β-fructo-
t it also hydrolyses sucrose.
in sucrose in the β-form.
optical rotation of the fruc-
...re for sucrose accounts for
all of the above facts:



Oxidation of sucrose with periodic acid confirms this structure (but not the nature of the glycosidic link). Three molecules of periodic acid are consumed, and one molecule of formic acid is produced. Subsequent oxidation with bromine water, followed by hydrolysis, gives glyoxylic, glyceric and hydroxypyruvic acids (Fleury *et al.*, 1942).

Beevers *et al.* (1947) examined sucrose sodium bromide dihydrate by X-ray analysis, and confirmed the stereochemical configuration found chemically, and also showed that the fructose ring is five-membered.

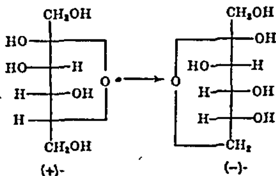
Sucrose has now been synthesised by Lemieux *et al.* (1953, 1956). 1:2-Anhydro-α-D-glucopyranose triacetate and 1:3:4:6-*tetra-O*-acetyl-D-fructofuranose were heated in a sealed tube at 100° for

```
 HC═══O─┐  CH₂OH                      HC═══O─┐  CH₂OH
  |     |    |                         |     |    |
 HCOH   └─── C ──┐                    CHO    └─── C ──┐
1 ---+---  |   |                        |           |    |
 HOCH   O   HOCH  |    3HIO₄ →  HCO₂H⁺   O          CHO  |
2 ---+---  |  3---+--  O                  |           |   O
 HCOH   |   HCOH  |                    CHO          CHO  |
  |     |    |    |                     |            |   |
 HC─────┘   HC────┘                    HC────┘       HC──┘
  |          |                          |            |
 CH₂OH      CH₂OH                      CH₂OH        CH₂OH

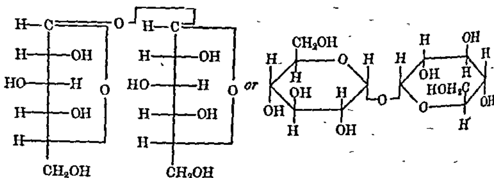
                                CHO          CH₂OH
                                 |             |
                                CO₂H          CO
 (i) Br₂/H₂O                                   |
 ──────────────→                 +      +     CO₂H
 (ii) hydrolysis                               +
                                CO₂H          CO₂H
                                 |             |
                                HCOH          HCOH
                                 |             |
                                CH₂OH         CH₂OH
```

104 hours. The product, sucrose octa-acetate, on deacetylation, gave sucrose. An interesting point about this synthesis is that it was anticipated by Lemieux (1953) on the basis of a conformational analysis of the properties of the anhydride.

One other point that is of interest is the "inversion" of sucrose on hydrolysis. Hydrolysis of sucrose gives first of all α-D(+)-glucopyranose and β-D(+)-fructofuranose (this is believed to be dextrorotatory), but the latter is unstable and immediately changes into the stable form, D(−)-fructopyranose (the rotation of (−)-fructose is much greater than that of (+)-glucose).



§14. Trehalose. This is believed to be α-D-glucopyranosyl-α-D-glucopyranoside. It is a non-reducing sugar which occurs in yeasts and fungi. It is hydrolysed by hydrochloric acid to two molecules of D-glucose; methylation of trehalose gives octa-*O*-methyltrehalose which, on hydrolysis, produces two molecules of 2·3·4·6-tetra-*O*-methyl-D-glucose (see §7a). The nature of the glycosidic link is uncertain, but there is some evidence to show that it is α:α, *e.g.*, the high positive rotation. Thus trehalose may be written:



§15. Maltose. This is 4-*O*-α-D-glucopyranosyl-D-glucopyranose. Maltose is hydrolysed by dilute acids to two molecules of D-glucose; it is a reducing sugar, undergoes mutarotation, and forms an osazone. Thus there is one free reducing group present, and since

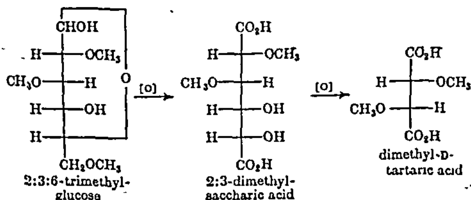
and this, on hydrolysis with very dilute cold hydrochloric acid, is converted into heptamethylmaltose, which has reducing properties. Thus the original octamethyl derivative must be methyl hepta-*O*-methyl-D-maltoside; this is further evidence that only *one* free reducing group is present in maltose. Hydrolysis of hepta-*O*-methylmaltose with moderately concentrated hydrochloric acid produces

D-glucose.

of the forn...

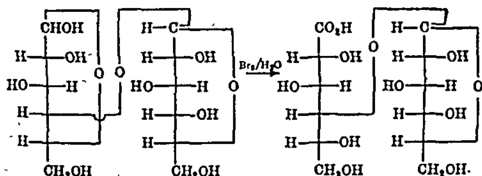
showed that it was a trimethyl derivative, and since it formed a phenylhydrazone but not an osazone, C_5 must therefore be attached to a methoxyl group. On further methylation, this trimethylglucose gave 2:3·4:6-tetra-*O*-methyl-D-glucose, and so the trimethyl compound must be one of the following: 2:3:4-, 2:3:6-, or 2:4:6-tri-*O*-methyl-D-glucose. Now, on careful oxidation with nitric acid the trimethylglucose forms a dimethylsaccharic acid. This acid contains two terminal carboxyl groups; one has been derived from

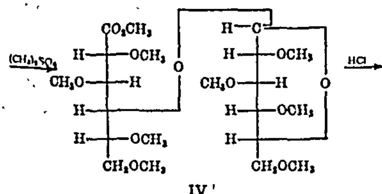
the free "aldehyde" group, and the other by oxidation at C_6, and since in its formation one methyl group is lost, this dimethylsaccharic acid must have been derived from a trimethylglucose having a methoxyl group at C_6. Thus the trimethylglucose must be either 2:3.6- or 2:4.6-tri-*O*-methyl-D-glucose. On further oxidation, the dimethylsaccharic acid forms dimethyl-D-tartaric acid; this can only arise from a precursor with two methoxyl groups on *adjacent* carbon atoms, and so it follows that the trimethylglucose must be 2:3:6-tri-*O*-methyl-D-glucose. This is confirmed by the fact that the other two possible compounds, *viz.*, 2.3:4- and 2.4:6-tri-*O*-methyl-D-glucose, have been synthesised, and were shown to be different from the trimethylglucose obtained from maltose. The foregoing reactions may thus be written:



From this it can be seen that structure I for maltose satisfies all the above facts. This structure, however, is not the only one that satisfies all the facts. The structure of the non-reducing half is certain, but that of the reducing half need not necessarily be pyranose as shown in I, since a furanose structure, II, would also give 2:3.6-tri-*O*-methyl-D-glucose. To decide whether C_4 (as in I) or C_5 (as in II) was involved in the glycosidic link, Haworth *et al.* (1926) oxidised maltose with bromine water to maltobionic acid, III, and this, on methylation, gave the methyl ester of octamethylmaltobionic acid, IV, which, on vigorous hydrolysis, gave 2:3.5.6-tetra-*O*-methyl-D-gluconic acid, V (as lactone), and 2.3:4:6-tetra-*O*-methyl-D-glucose, VI. V can be obtained only if maltose has structure I; structure II would have given 2:3:4:6-tetra-*O*- [illegible]

is a reducing sugar, C_1 (of the reducing half) is free, and since



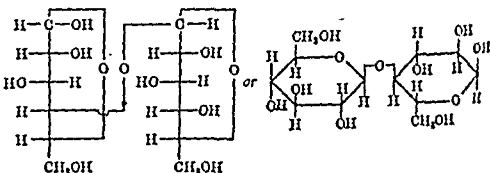


CO₂H / H—OCH₃ / CH₃O—H / H—OH / H—OCH₃ / CH₂OCH₃ + CHOH / H—OCH₃ / CH₃O—H / H—OCH₃ / H / CH₂OCH₃ / O

V VI

§16. **Celloblose** (4-*O*-β-D-glucopyranosyl-D-glucopyranose). Cellobiose is hydrolysed by dilute acids to two molecules of D(+)-glucose; since this hydrolysis is also effected by emulsin, the glycosidic link must be β. Cellobiose is a reducing sugar, and so one reducing group is free. Methylation, followed by hydrolysis, gives 2:3:6-trimethyl-D-glucose and 2:3:4:6-tetramethyl-D-glucose (these

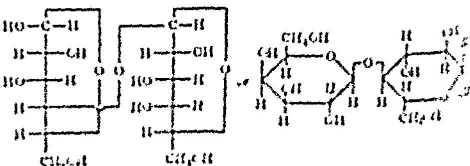
are the same products obtained from maltose, §15). Oxidation with bromine water converts cellobiose into cellobionic acid, and this, on methylation followed by hydrolysis, gives 2:3.5.6-tetramethylgluconic acid and 2:3:4:6-tetramethylglucose (again the same products as for maltose). Thus cellobiose and maltose differ only in that the former has a β-glycosidic link, whereas the latter has an α-. Thus cellobiose is (α-form):



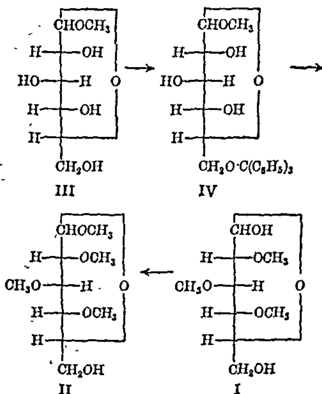
Degradation experiments confirm the C_4 linkage (see also §7g), and the structure has also been confirmed by synthesis (*e.g.*, Stacey, 1946).

§17. Lactose (4-*O*-β-D-galactopyranosyl-D-glucopyranose). Lactose is a reducing sugar, and is hydrolysed by dilute acids to one molecule of D(+)-glucose and one molecule of D(+)-galactose. Since lactose is hydrolysed by lactase (which has been shown to be identical with the β-glycosidase in emulsin), the two monosaccharide

on vigorous hydrolysis, gives 2:3:6-tri-*O*-methyl-D-glucose (see §12) and 2:3:4:6-tetra-*O*-methyl-D-galactose; thus glucose is the reducing half. Oxidation with bromine water converts lactose into lactobionic acid, and this, on methylation followed by hydrolysis, gives 2:3:5:6-tetra-*O*-methyl-D-gluconic acid and 2:3:4:6-tetra-*O*-methyl-D-galactose. Lactose is therefore (β-form) (see also §7g):

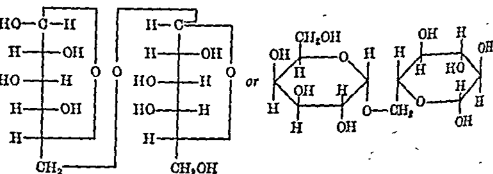


§18. Melibiose (6-*O*-α-D-galactopyranosyl-D-glucopyranose) This disaccharide is obtained from the trisaccharide, raffinose (§20);



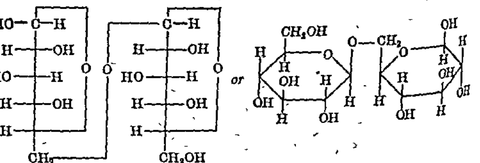
it is a reducing sugar, forms an osazone, and undergoes mutarotation. When hydrolysed by dilute acids, melibiose gives D-glucose and D-galactose. Methylation converts melibiose into methyl heptamethylmelibioside, and this, on hydrolysis, forms 2:3:4-trimethyl-D-glucose and 2.3.4:6-tetramethyl-D-galactose. The structure of the former has been established as follows. The trimethylglucose, I, readily forms a crystalline methyl trimethylglucoside, II. Now methyl glucopyranoside, III, can be converted into the 6-trityl derivative, IV (see §9), and this, on methylation followed by removal of the trityl group, gives II. Thus II must be methyl 2:3:4-tri-*O*-methyl-D-glucopyranoside, and consequently I is 2.3:4-tri-*O*-methyl-D-glucose. From the foregoing facts, it can be seen that galactose is the non-reducing half of melibiose, and that its reducing group is linked to C_6 of glucose, the reducing half. This has been confirmed by oxidation of melibiose with bromine water to melibionic acid, and this, on methylation followed by hydrolysis, gives 2:3:4.5-tetra-*O*-methyl-D-gluconic acid and 2:3:4.6-tetra-*O*-methyl-D-galactose; the structure of the former is shown by the

fact that, on oxidation with nitric acid, it forms tetramethylsaccharic acid. There has been some doubt about the nature of the glycosidic link, but the evidence appears to be strongly in favour of α-. Thus the structure of melibiose is (β-form) [see also §7g]:



Melibiose has been synthesised chemically.

§19. **Gentiobiose** (6-*O*-β-D-glucopyranosyl-D-glucopyranose). This was originally obtained from the trisaccharide, gentianose (§20), but it also occurs in some glycosides, *e.g.*, amygdalin (§27). Gentiobiose is a reducing sugar, forms an osazone and undergoes mutarotation; hydrolysis with dilute acids produces two molecules of D-glucose. Since this hydrolysis is also effected by emulsin, the glycosidic link must be β-. Methylation, followed by hydrolysis, gives 2:3:4-trimethyl-D-glucose and 2:3:4:6-tetramethyl-D-glucose. Oxidation to gentiobionic acid, this then methylated and followed by hydrolysis, gives 2:3:4:5-tetramethyl-D-gluconic acid and 2:3:4:6-tetramethyl-D-glucose. Thus gentiobiose is (β-form):



Gentiobiose has been synthesised chemically.

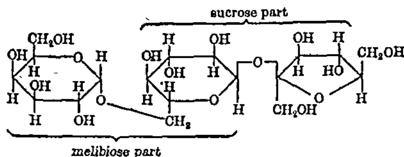
§20. **Trisaccharides.** The trihexose trisaccharides have the molecular formula $C_{18}H_{32}O_{16}$. They are of two types, reducing and non-reducing. **Manninotriose** is the only reducing trisaccharide that has been isolated from natural sources. All the others of this

group have been obtained by degrading polysaccharides or by synthesis, *e.g.*, **cellotriose** from cellulose. Two important non-reducing trisaccharides are raffinose and gentianose.

Raffinose occurs in many plants, particularly beet. Controlled hydrolysis with dilute acids gives D-fructose and melibiose; vigorous hydrolysis gives D-fructose, D-glucose and D-galactose. It is also hydrolysed by the enzyme invertase to fructose and melibiose, and by an α-glycosidase to galactose and sucrose. These facts show that the three monosaccharide molecules are linked in the following order:

galactose—glucose—fructose

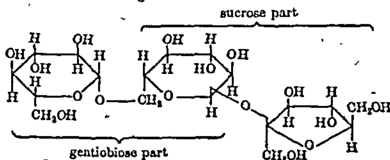
This arrangement is confirmed by the products obtained by methylation of raffinose, followed by hydrolysis, *viz.*, 2:3·4:6-tetramethylgalactose, 2:3:4-trimethylglucose, and 1:3:4.6-tetramethylfructose. Furthermore, since the structures of sucrose (§13) and melibiose (§18) are known, the structure of raffinose must therefore be:



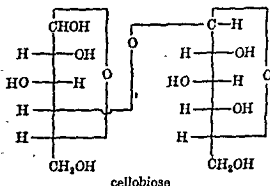
Gentianose occurs in gentian roots. Controlled hydrolysis with dilute acids gives D-fructose and gentiobiose; this hydrolysis is also effected by the enzyme invertase. Emulsin also hydrolyses gentianose to D-glucose and sucrose. Thus the arrangement of the three monosaccharide molecules is:

glucose—glucose—fructose

Hence the structure of gentianose is:



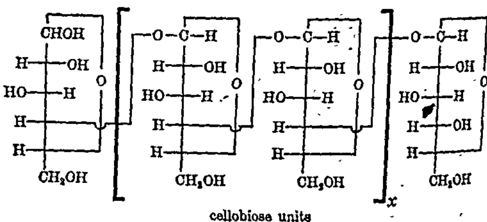
and concentrated sulphuric acid), cellulose forms cellobiose octa-acetate. Thus the cellobiose unit is present in cellulose, and since the structure of cellobiose is known (see §16), it therefore follows that the glucose units are present in the pyranose form, *i.e.*, C_5 is involved in ring formation, and so the glucose units are linked C_1—C_4. The isolation of cellobiose indicates also that *pairs* of glucose units are joined by β-links, but it does indicate whether the



cellobiose

links between the glucose units are the same (all β-) or alternate (α and β), since all the links could be β-, or each pair of cellobiose
iot
of
a
vn
ay
conclude that *all* the links in cellulose are β-. This conclusion is supported by other evidence, *e.g.*, the kinetics of hydrolysis of cellulose.

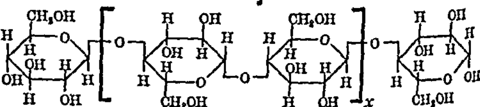
Cellulose forms colloidal solutions in solvents in which glucose is soluble, and so it is inferre
Moreover, since cellulose fo
that the molecule is linear;
nature of the molecule, and that the cellulose molecule has a long length. Hence a possible structure for cellulose is:



cellobiose units

Ia

or



Ib

It should be noted that in the structure given for cellulose, the first glucose unit in Ia (*i.e.*, the one on the left-hand side; this unit is on the right-hand side in Ib) has a free reducing group, but since this group is at the end of a very long chain, its properties tend to be masked; thus cellulose does not exhibit the strong reducing properties of the sugars.

The cellulose molecule is not planar, but has a screw-axis, each glucose unit being at right angles to the previous one. Although free rotation about the C—O—C link might appear possible at first sight, it apparently does not occur owing to the spatial effect (steric effect). This and the close packing of the atoms give rise to a rigid chain molecule. The long chains are held together by hydrogen bonding, and thus cellulose has a three-dimensional brickwork. This would produce strong fibres with great rigidity but no flexibility, and consequently, although the fibres would have great tensile strength, they could not be knotted without snapping. Since the fibres can be knotted without snapping, they must possess flexibility, and the presence of the latter appears to be due to the partly amorphous character of cellulose.

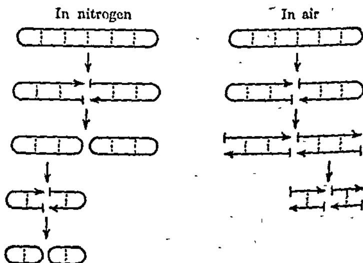
The chemical structure of cellulose appears to be more complicated than the one given above. Schmidt *et al.* (1932) showed

that carboxyl groups are present in carefully purified cotton fibres. Kleinert *et al.* (1944) have suggested that various other groups, which are not necessarily carbohydrate in nature, may bind the glucose chains together. It should be remembered, in this connection, that 100 per cent. glucose is never obtained from the hydrolysis of cellulose.

The molecular weight of cellulose. Owing to its insolubility, simple methods of molecular weight determination (depression of freezing point and elevation of boiling point) cannot be applied to cellulose.

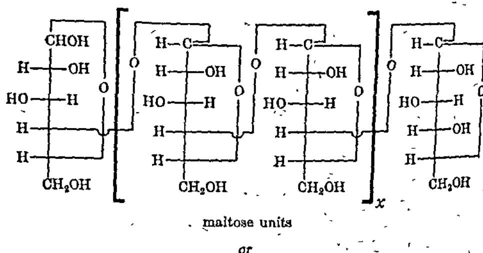
Chemical methods. Examination of the formula of cellulose shows that on methylation, followed by hydrolysis, the end unit (the non-reducing end) would contain four methoxyl groups, and all the other units three. Hence, by the determination of the percentage of the tetramethyl derivative (2:3.4.6-) it is therefore possible to estimate the length of the chain. Haworth (1932) separated the methylated glucoses by vacuum distillation; Hibbert (1942) used fractional distillation; Bell (1944), using silica, and Jones (1944), using alumina, effected separation by means of chromatography. The value for the molecular weight of cellulose was found to be between 20,000 and 40,000 (Haworth, 1932); this corresponds approximately, to 100 to 200 glucose units. This "end-group assay", however, gives rise to the following difficulty. When cellulose is very carefully prepared from cotton, and then methylated in an atmosphere of nitrogen, *i.e.*, in the *absence* of oxygen, no 2:3:4:6-tetramethylglucose was obtained after hydrolysis (Haworth *et al.*, 1939). One explanation that has been offered is that during methylation under ordinary conditions, *i.e.*, in air, cellulose is partially degraded, *e.g.*, osmotic pressure measurements carried out on methylated cellulose, produced by two methylations in air, gave a value of 190 glucose units; sixteen successive methylations in air gave a methylated cellulose of 45 glucose units, as estimated by osmotic pressure measurements (Haworth *et al.*, 1939). Haworth explained these results by suggesting that the cellulose molecule consists of a very large loop, which undergoes progressive shortening on methylation. When the methylation is carried out in an atmosphere of nitrogen, the exposed ends of the shortened loop recombine, but cannot do so when methylated in the presence of air. Haworth also suggested that in order that the two chains should be held parallel to form a loop, it is necessary to have cross-linkages holding the two sides together. The nature of these suggested cross-links is unknown. If primary valencies were involved, then some *dimethylglucose* should be obtained from the hydrolysate. Some of this compound has in fact been isolated

but it is not certain that it is actually present in the methylated cellulose, since it may arise by demethylation during the degradation of the methylated cellulose. The following is a pictorial representation of Haworth's suggestion.



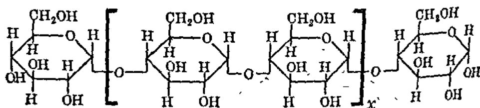
By means of chromatography, McGilvray (1953) has detected 2:3:4:6-tetra-*O*-methyl-D-glucose in the hydrolysate after the methylation of cellulose *in an atmosphere of nitrogen*. Thus degradation of the chain has occurred under these conditions, and so there is no evidence for the linking of the end groups in the absence of oxygen. Furthermore, McGilvray determined the degree of polymerisation from viscosity and osmotic pressure measurements, and also from the end-group assay. The values obtained from the first two methods were greater than that obtained from the third method, and McGilvray suggests these results may be accounted for by assuming a slightly branched structure for the soluble methylcelluloses.

A number of other chemical methods have been used for estimating the molecular weight of cellulose, *e.g.*, that of Hirst *et al.* (1945); this is based on the periodate oxidation (§7g). Examination of the formula of cellulose shows that the terminal reducing unit would give two molecules of formic acid and one of formaldehyde (this reducing unit, which is left in I*a*, behaves as the open-chain molecule, since it is *not* a glycoside), whereas the other terminal unit (right in I*a*) would give one molecule of formic acid; *i.e.*, one cellulose molecule gives three molecules of formic acid and one of formaldehyde. Estimation of the formic acid produced gives the value of the chain-length as approximately 1,000 glucose



maltose units

or



The Haworth end-group assay (1932) showed that starch is composed of approximately 24–30 glucose units. Thus starch is a linear molecule, at least as far as 24–30 units. Haworth, however, pointed out that this was a *minimum* chain-length, and that starches may differ by having different numbers of this repeating unit (see also below). Viscosity measurements, however, showed the presence of a highly branched structure. Now, it has long been known that starch can be separated into two fractions, but it is only fairly recently that this separation has been satisfactorily carried out; the two fractions are α-amylose (the A-fraction; 17–34 per cent.) and β-amylose (amylopectin, or the B-fraction). The fractionation has been carried out in several ways, *e.g.*, *n*-butanol is added to a hot colloidal solution (aqueous) of starch, and the mixture allowed to cool to room temperature. The A-fraction is precipitated, and the B-fraction is obtained from the mother liquors by the addition of methanol (Schoch, 1942). Haworth *et al.* (1946) have used thymol to bring about selective precipitation.

α-Amylose is soluble in water, and the solution gives a blue colour with iodine. β-Amylose is insoluble in water, and gives a violet colour with iodine. Both amyloses are mixtures of polymers,

end-group assay by methylation shows the presence of one unit with four free hydroxyl groups per 24–30 glucose units; the same results are also obtained by the periodate method. Thus the 3,000 units are joined in such a manner as to give about 100 end units; it therefore follows that the chain must be *branched*. The problem is further complicated by the fact that Hirst (1940), after methylating amylopectin and hydrolysing the product, obtained, in addition to tri- and tetra-*O*-methyl-D-glucose, about 3 per cent. of 2 3-di-*O*-methyl-D-glucose. This has been taken to mean that some glucose units are also joined by C_1 and C_6 atoms. Furthermore, in certain experiments, enzymic hydrolysis has given a small amount of 1 6 α-linked diglucose, *i e.*, *iso*maltose is also present in amylopectin (Montgomery *et al.*, 1947, 1949). Wolfrom *et al.* (1955, 1956) have obtained evidence that there is also an α-D-1,3-bond in amylopectin; the principal bond is α-D-1:4, and branching occurs through α-D-1:6-bonds.

The branching of the chains in amylopectin is supported by the following evidence:

(i) Amylopectin acetate does not form fibres; fibre formation is characteristic of *linear* molecules

(ii) Diastase (β-amylase) hydrolyses amylopectin to give only about 50 per cent. of maltose. Thus there are "blocked" points, and these will occur at the branch points.

(iii) Amylopectin solutions do not show an orientation of the molecules in the direction of flow in the concentric cylinder technique; the molecules are therefore not linear.

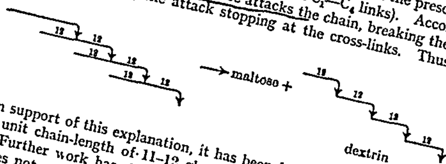
The detailed structure of amylopectin is still not settled. Haworth and Hirst (1937) suggested a laminated formula for amylopectin; each line represents a basal chain of 24–30 glucose units joined by α C_1—C_4 links, and each arrow head represents the

joining of the terminal reducing group (C_1) of each chain to the central glucose member (at C_6) of the next chain If it is branched in the above fashion, then methylated amylopectin should give some dimethylglucose on hydrolysis. Since 2 3-di-*O*-methyl-D-glucose is actually obtained, the link must be C_1 of one chain to C_6 of the next If the unions are as regular as this, then there will be one C_1—C_6 link for every one end group Hirst *et al.* (1947), however, showed by the end-group assay by the periodic a[illegible]

method that amylopectin contains only traces of glucose residues joined solely by C_1—C_4 links.

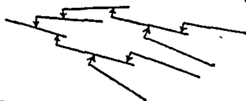
Prolonged methylation of amylopectin produces a diminution of the molecular size (as determined by physical methods); *e.g.*, methylation of starch seventeen times changed the particle size from 3,000 glucose units to 190 units (Averill, 1939). This diminution in particle size cannot be due to the break-down of the basal chains, since the end-group assay always gives the same basal chain-length, whether the methylation is carried out in air or in an atmosphere of nitrogen. Haworth therefore suggested that this diminution in particle size is due to the "disaggregation" of the basal chains.

As pointed out previously, β-amylase gives only 50 per cent. of maltose with amylopectin. The high molecular weight residue is known as **dextrin**, and this is not degraded because of the presence of the C_1—C_6 link (β-amylase breaks only α C_1—C_4 links). According to Haworth (1946), β-amylase attacks the chain, breaking them into units of two, the attack stopping at the cross-links. Thus:



n support of this explanation, it has been found that dextrin has unit chain-length of 11–12 glucose units.

Further work has shown that the Haworth laminated formula does not satisfy the behaviour of amylases on amylopectin; the formula is far too regular (*cf.* Hirst's work, above). Meyer (1940) proposed a highly branched structure; this fits the behaviour of



the amylases better. Furthermore, mathematical calculations have shown that the regular form is unlikely. A difficulty of the Meyer structure, however, is that amylopectin would be globular; this is not in keeping with all the evidence.

§23. Some other polysaccharides. A number of other polysaccharides besides cellulose and starch also occur naturally, and some of these are described briefly below.

Glycogen. This is the principal reserve carbohydrate in animals. It is hydrolysed by β-amylase to maltose, and molecular weight determinations by physical methods give values between 1 and 2,000,000.

Inulin. This is a fructosan, and occurs in dahlia tubers, dandelion roots, etc. Acid hydrolysis gives D-fructose, but if inulin is first methylated and then hydrolysed, 3:4:6-tri-*O*-methyl-D-fructose is the main product, thus indicating that inulin is composed of fructofuranose units.

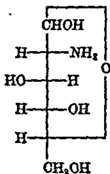
Mannans are polysaccharides which yield only mannose on hydrolysis; they are found in ivory nut, seaweeds, bakers' yeast, etc. Similarly, **galactans** yield only galactose on hydrolysis; they occur in seeds, wood, etc. There are also polysaccharides which contain pentose residues only, *viz.* **pentosans,** *e.g.*, **xylans** give D-xylose; *arabans* give L-arabinose. Some pentosans are composed of both xylose and arabinose, and other polysaccharides are composed of pentose and hexose units, *e.g.*, *xylo-glucans* (xylose and glucose), *arabo-galactans*, etc. In addition to these neutral polysaccharides, there are also the acid polysaccharides. These are gums and mucilages, and owe their acidity to the presence of uronic acids. Gums are substances which swell in water to form gels (or viscous solutions), *e.g.*, gum arabic and gum tragacanth; on hydrolysis, the former gives arabinose, galactose, rhamnose and glucuronic acid, and the latter xylose, L-fucose and galacturonic acid. Mucilages are polysaccharides which swell in water to form viscous solutions; on hydrolysis, they give galacturonic acid, arabinose, xylose, etc. The **hemi-celluloses** (which are widely distributed in the cell-wall of plants) also contain both uronic acids (glucuronic or galacturonic) and pentoses (xylose, arabinose).

Pectin. This occurs in plants, particularly fruit juices. It is composed of D-galacturonic acid residues and the methyl ester.

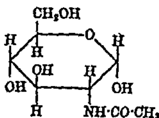
Alginic acid. This occurs in the free state and as the calcium salt in various seaweeds. Hydrolysis of alginic acid produces D-mannuronic acid.

Chitin. This is the polysaccharide that is found in the shells of crustaceans. Hydrolysis of chitin by acids produces acetic acid and D-glucosamine (chitosamine; 2-aminoglucose). Chitin is also

replaces glucose).



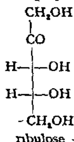
D-glucosamine



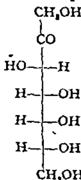
N-acetylglucosamine

N-Methyl-L-glucosamine is a component of streptomycin (see §7. XVIII).

§23a. Photosynthesis of carbohydrates. The scheme outlined below is largely that proposed by Calvin *et al.* (1954). These authors exposed certain algae to carbon dioxide (labelled with ^{14}C) and light, then killed the algae and extracted with ethanol and chromatographed (on paper) the extract. Two monosaccharides, ribulose and sedoheptulose, play an essential part in the photosynthesis of



ribulose



sedoheptulose

carbohydrates, and the steps involved are as follows:

(i) Ribulose diphosphate accepts one molecule of carbon dioxide and one of water.

(ii) The product now splits into two molecules of phosphoglyceric acid ($CH_2O{\cdot}PO_3H_2{\cdot}CHOH{\cdot}CO_2H$).

(iii) Phosphoglyceric acid undergoes reduction to phosphoglyceraldehyde.

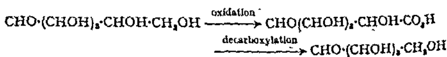
(iv) Two molecules of phosphoglyceraldehyde combine to form hexose phosphate.

(*va*) Hexose phosphate forms disaccharides and polysaccharides.

(*vb*) A molecule of hexose phosphate reacts with a molecule of

phosphoglyceraldehyde to form ribulose phosphate and a tetrose phosphate. The latter reacts with a molecule of phosphoglyceraldehyde to produce sedoheptulose phosphate which, in turn, also reacts with a molecule of phosphoglyceraldehyde to produce one molecule of ribose phosphate and one molecule of ribulose phosphate. The ribose phosphate is then converted into ribulose phosphate, thus completing the cycle.

All the aldohexoses and all the aldopentoses are interconvertible by inversion of one asymmetric carbon atom, but how this occurs in the plant is not certain. Furthermore, aldohexoses may be stepped down to aldopentoses, and again how this occurs is not certain; one suggestion is (see also §32a. VIII):



$$CHO\cdot(CHOH)_3\cdot CHOH\cdot CH_2OH \xrightarrow{\text{oxidation}} CHO(CHOH)_3\cdot CHOH\cdot CO_2H$$
$$\xrightarrow{\text{decarboxylation}} CHO\cdot(CHOH)_3\cdot CH_2OH$$

GLYCOSIDES

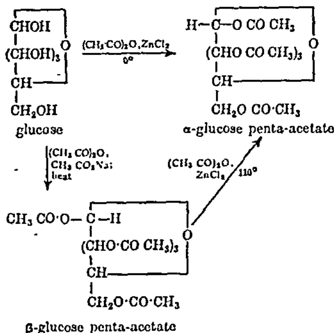
§24. Introduction. A great variety of glycosides occur in plants. The simple glycosides are colourless, soluble in water, and are optically active; they do not reduce Fehling's solution. On

of the plant. Maceration of the plant thus produces hydrolysis of the glycoside by bringing the enzyme in contact with the glycoside. Glucose has been found to be the most common sugar component; when methylated and then hydrolysed, most glycosides give 2:3:4:6-tetra-*O*-methyl-D-glucose. Thus most glycosides are β-D-glucopyranosides.

Synthesis of glycosides. The synthesis of a glycoside uses an acetobromohexose as the starting material; this compound is now named systematically as a tetra-*O*-acetyl-D-hexopyranosyl 1-bromide, *e.g.*, if the hexose is glucose, then the α-form will be tetra-*O*-acetyl-α-D-glucopyranosyl 1-bromide.

When glucose is treated with acetic anhydride at 0° in the presence of zinc chloride, the product is 1:2:3:4:6-penta-*O*-acetyl-α-D-glucose (α-D-glucose penta-acetate). If, however, glucose is heated with acetic anhydride in the presence of sodium acetate, the product is 1:2:3:4:6-penta-*O*-acetyl-β-D-glucose. Furthermore,

the β-isomer may be converted into the α- by heating with acetic anhydride at 110° in the presence of zinc chloride.



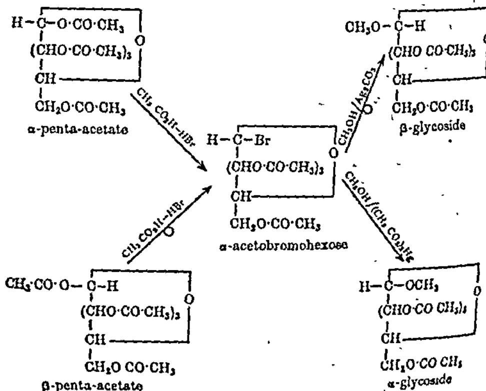
β-glucose penta-acetate

These penta-acetates are readily hydrolysed to glucose by means of dilute aqueous sodium hydroxide, ethanolic ammonia at 0°, or by methanol containing a small amount of sodium methoxide. When dissolved in glacial acetic acid saturated with hydrogen bromide, the glycosidic acetoxyl group of a hexose penta-acetate is replaced by bromine to give an α-acetobromohexose; the α-isomer is obtained whether the penta-acetate used is the α- or β-compound (Fischer, 1911). Thus a Walden inversion occurs with the β-compound (§1. III).

Scheurer *et al.* (1954) have synthesised acetobromo sugars in good yield as follows. Bromine is added to a suspension of red phosphorus in glacial acetic acid, and to this solution (which now contains acetyl bromide) is added the sugar or acetylated sugar, the latter giving the better yields.

The bromine atom in these acetobromohexoses is very reactive. Thus it may be replaced by a hydroxyl group when the acetobromohexose is treated with silver carbonate in moist ether (Fischer *et al.*, 1909), or by an alkoxyl group when treated with an alcohol in the presence of silver carbonate (Königs and Knorr, 1901). *In* the latter reaction the yields are improved if anhydrous calcium sulphate and a small amount of iodine are used instead of silver

carbonate (Evans *et al.*, 1938). In either case, the α-acetobromohexose gives the β-glycoside. On the other hand, if mercuric acetate is used instead of silver carbonate, then the α-glycoside is obtained. The foregoing reactions may thus be written (using the symbol ⟶⊙⟶ to represent a Walden inversion; see §1. III).



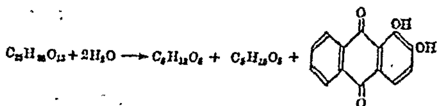
§25. **Indican.** This glycoside occurs in the leaves of the indigo plant and in the woad plant. When the leaves are macerated with water, the enzyme present hydrolyses indican to glucose and indoxyl, and the latter, on exposure to air, is converted into indigotin (see Vol. I).

The molecular formula of indican is $C_{14}H_{17}O_6N$, and since it gives D-glucose and indoxyl on hydrolysis, it is therefore indoxyl D-glucoside. When indican is methylated (with methyl iodide in the presence of dry silver oxide), tetramethylindican is obtained, and this, on hydrolysis with methanol containing 1 per cent. hydrogen chloride, gives indoxyl and methyl 2,3,4,6-tetra-*O*-methyl-D-glucoside. Thus the glucose molecule is present in the pyranose form, and since indican is hydrolysed by emulsin, the glycosidic link must be *β*. Thus the structure of indican is III.

and this has been confirmed by synthesis from indoxyl, I, and *tetra-O-acetyl-α-D-glucopyranosyl 1-bromide*, II, as follows:

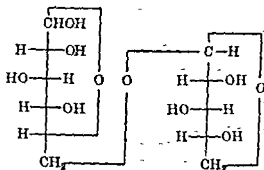
I + II $\xrightarrow[Q]{Ag_2CO_3}$ (tetra-O-acetyl glucoside) $\xrightarrow{NH_3 / C_2H_5OH}$ III

§26. Ruberythric acid. This occurs in the madder root, and on hydrolysis, it was originally believed to give one molecule of alizarin and two molecules of D-glucose. Jones and Robertson (1933), however, showed that two molecules of D-glucose were not present in the hydrolysate; a mixture of two sugars was actually present, D-glucose and D-xylose. Thus the molecular formula of ruberythric acid is $C_{25}H_{26}O_{13}$, and not, as was originally believed, $C_{26}H_{28}O_{14}$. Thus the hydrolysis is:



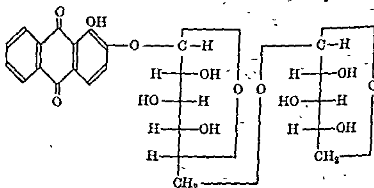
Jones and Robertson also showed that the two monosaccharide molecules were present in the form of the disaccharide primeverose.

Now, this disaccharide is 6-*O*-β-D-xylopyranosyl-D-glucopyranose (Helferich, 1927), and it therefore follows that alizarin is linked to the



primeverose

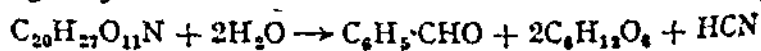
glucose half of the primeverose molecule. Further work has shown that the glucosidic link is β, and that it is the 2-hydroxyl group of alizarin that is involved. Thus the structure of ruberythric acid is:



§27. Amygdalin. This occurs in bitter almonds. The molecular formula is $C_{20}H_{27}O_{11}N$, and it is hydrolysed by acids to one molecule of benzaldehyde, two molecules of D-glucose, and one of hydrogen cyanide.

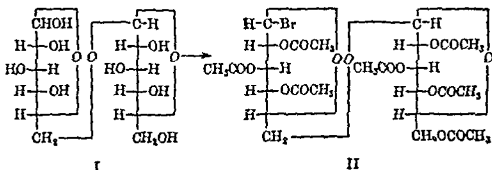
$$C_{20}H_{27}O_{11}N + 2H_2O \rightarrow C_6H_5{\cdot}CHO + 2C_6H_{12}O_6 + HCN$$

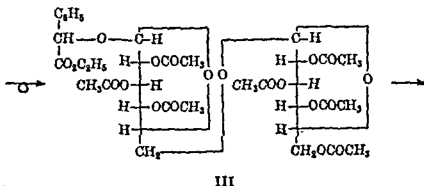
Since emulsin also brings about this hydrolysis, amygdalin must contain a β-glycosidic link. On the other hand, the enzyme zymase hydrolyses amygdalin into one molecule of glucose and a glucoside of (+)-mandelonitrile (this compound is

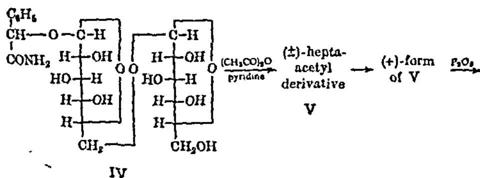


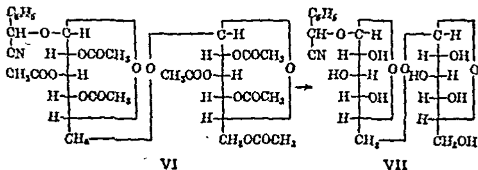
$$C_{20}H_{27}O_{11}N + H_2O \rightarrow C_6H_{12}O_6 + C_6H_5{\cdot}CH(CN){\cdot}O{\cdot}C_6H_{11}O_5$$

identical with *prunasin*, a naturally occurring glucoside). Thus the aglycon of amygdalin is (+)-mandelonitrile, and the sugar is a disaccharide. Haworth *et al.* (1922, 1923) have shown that this disaccharide is gentiobiose (§19), and have synthesised amygdalin


I
II


III

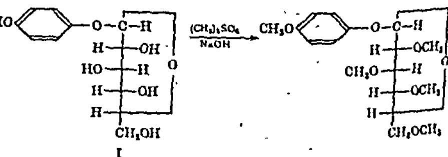

IV
(CH₃CO)₂O
pyridine
(±)-hepta-acetyl derivative
V
(+)-form of V
P₂O₅

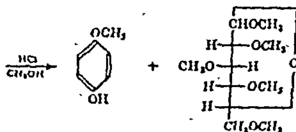

VI
VII

(in 1924) as follows. Gentiobiose, I, was converted into hepta-acetyl-bromogentiobiose, II, by means of acetic anhydride saturated with hydrogen bromide, and then II was condensed with racemic ethyl mandelate in the presence of silver oxide, whereby the β-glycoside, III, was obtained. Treatment of this with ethanolic ammonia hydrolysed the acetyl groups, and at the same time converted the ester group into the corresponding amide; thus the ($\pm$)-amido-glycoside, IV, was obtained. IV was then treated with acetic anhydride in pyridine solution, and the ($\pm$)-hepta-acetyl derivative of the amide, V, was then separated into its diastereoisomers by fractional crystallisation (the mandelic acid portion is + and −, the gentiobiose portion is +; hence the two forms present are ++ and −+, *i.e.*, they are diastereoisomers). The (+)-form was then dehydrated with phosphorus pentoxide to give the (+)-nitrile, VI, and this, on de-acetylation with ethanolic ammonia, gave (+)-amygdalin, VII, which was shown to be identical with the natural compound.

§28. Arbutin and Methylarbutin. **Arbutin** is hydrolysed by emulsin to give one molecule of D-glucose and one of quinol; thus arbutin is a β-glucoside. When methylated (with methyl sulphate

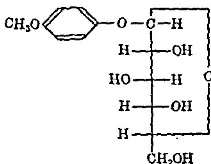
quinol (Macbeth *et al.*, 1923); structure I for arbutin accounts for all these facts.





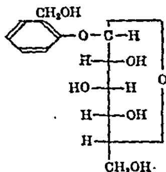
Pentamethylarbutin has been synthesised by converting 2:3:4:6-tetra-*O*-methyl-D-glucose into tetra-*O*-methyl-α-D-glucopyranosyl 1-bromide, and condensing this with monomethylquinol; the product is identical with the methylated natural compound.

Methylarbutin. This is hydrolysed by emulsin to one molecule of D-glucose and one molecule of monomethylquinol; thus methylarbutin is a β-glucoside, and its structure is:



Methylarbutin has been synthesised by condensing tetra-*O*-acetyl-α-D-glucopyranosyl 1-bromide with monomethylquinol in the presence of silver carbonate, followed by de-acetylation.

§29. Salicin. This is hydrolysed by emulsin to one molecule of D-glucose and one of salicyl alcohol (saligenin). Thus salicin is a β-glucoside, but it is not possible to tell from the hydrolytic products whether it is the phenolic or alcoholic group of the salicyl alcohol which forms the glycosidic link. Which group is involved is readily shown as follows (Irvine *et al.*, 1906). Oxidation of salicin with nitric acid forms *helicin*, and this, on hydrolysis, gives glucose and salicylaldehyde. Thus the phenolic group in salicyl alcohol must form the glucoside. Methylation of salicin produces pentamethylsalicin, and this, on hydrolysis, gives 2:3:4:6-tetra-*O*-methyl-D-glucose. Hence the glucose residue is in the pyranose



form; the structure given for salicin fits the foregoing facts. This structure has been confirmed by condensing tetra-*O*-methyl-α-D-glucopyranosyl 1-bromide with salicyl alcohol, and then methylating the product. The pentamethylsalicin so obtained was identical with the methylated natural product (Irvine *et al.*, 1906).

§30. Sinigrin. This glycoside occurs in black mustard seed, and on hydrolysis with the enzyme myrosin, D-glucose, allyl *iso*thiocyanate and potassium hydrogen sulphate are obtained.

$$C_{10}H_{16}O_9NS_2K + H_2O \rightarrow C_6H_{12}O_6 + CH_2{=}CH{\cdot}CH_2{\cdot}NCS + KHSO_4$$

Sodium [illegible]
obtained [illegible]
the gluco[illegible]
still not certain whether the glycosidic link is α or β. The following is a possible structure for sinigrin:

```
 CH═══════┐──S─C═N·CH₂·CH═CH₂
 │        │    │
H─┼─OH    │    O·SO₃K
 │        │
HO─┼─H    O
 │        │
H─┼─OH    │
 │        │
H─┼───────┘
 │
 CH₂OH
```

READING REFERENCES

Rosanoff, On Fischer's Classification of Stereoisomers, *J. Amer. Chem. Soc.*, 1906, **28**, 114.

Rules of Carbohydrate Nomenclature, *J.C.S.*, 1952, 5108.

Haworth, *The Constitution of Sugars*, Arnold (1929).

Honeyman, *Chemistry of the Carbohydrates*, Oxford Press (1948).

Percival, *Structural Carbohydrate Chemistry*, Muller (1950).

Pigman and Goepp, *Chemistry of the Carbohydrates*, Academic Press (1948).

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. (i) Vol. II (1943, 2nd ed.). Ch. 20, 21. Carbohydrates. Ch. 22. Cellulose. (ii) Vol. IV (1953). Ch. 9. Starch.

Ann. Reports (*Chem. Soc.*). (i) 1946, **43**, 30. The Molecular Weight and Dimensions of Macromolecules in Solution. (ii) 1953, **50**, 253, 288. Carbohydrates.

Percival, Carbohydrate Sulphates, *Quart. Reviews* (*Chem. Soc.*), 1949, **3**, 369.

Barker and Bourne, Enzymic Synthesis of Polysaccharides, *Quart. Reviews* (*Chem. Soc.*), 1953, 7, 56.
Hudson, Emil Fischer's Discovery of the Configuration of Glucose, *J. Chem. Educ.*, 1941, 18, 353.
Hirst, Obituary Notice of Haworth, *J.C.S.*, 1951, 2790.
Haworth, Starch, *J.C.S.*, 1946, 543.
Hirst, The Occurrence and Significance of the Pentose Sugars in Nature, and their Relationship to the Hexoses, *J.C.S.*, 1949, 522.
Advances in Carbohydrate Chemistry, Academic Press (1945–)
McGilvray, The Detection and Determination of "End Groups" in O-Methylcellulose, *J.C.S.*, 1953, 2577.
Head and Hughes, The Oxidation of Cellobiose by Periodate, *J.C.S.*, 1954, 603.
Reeves, The Shape of Pyranoside Rings, *J. Amer. Chem. Soc.*, 1950, 72, 1499.
Manners, The Enzymic Degradation of Polysaccharides, *Quart. Reviews* (*Chem. Soc.*), 1955, 9, 73.
Hirst, The Hudson Memorial Lecture, *J.C.S.*, 1954, 4042.
Sir Robert Robinson, *The Structural Relationships of Natural Products*, Oxford Press (1955).
Downes, *The Chemistry of Living Cells*, Longmans, Green (1955).
Lemieux and Huber, A Chemical Synthesis of Sucrose, *J. Amer. Chem. Soc.*, 1956, 78, 4117.

CHAPTER VIII

TERPENES

§1. Introduction. The terpenes form a group of compounds the majority of which occur in the plant kingdom; a few terpenes have been obtained from other sources. The simpler mono- and sesqui-terpenes are the chief constituents of the essential oils; these are the volatile oils obtained from the sap and tissues of certain plants and trees. The essential oils have been used in perfumery from the earliest times. The di- and tri-terpenes, which are not steam volatile, are obtained from plant and tree gums and resins. The tetraterpenes form a group of compounds known as the *carotenoids*, and it is usual to treat these as a separate group (see Ch. IX). Rubber is the most important polyterpene.

Most natural terpene hydrocarbons have the molecular formula $(C_5H_8)_n$, and the value of n is used as a basis of classification. Thus we have the following classes (these have already been mentioned above):

(i) Monoterpenes, $C_{10}H_{16}$. (ii) Sesquiterpenes, $C_{15}H_{24}$.
(iii) Diterpenes, $C_{20}H_{32}$. (iv) Triterpenes, $C_{30}H_{48}$.
(v) Tetraterpenes, $C_{40}H_{64}$ (these are the carotenoids).
(vi) Polyterpenes, $(C_5H_8)_n$.

In addition to the terpene hydrocarbons, there are the oxygenated derivatives of each class which also occur naturally, and these are mainly alcohols, aldehydes or ketones.

The term *terpene* was originally reserved for those hydrocarbons ... he term now ... is, however, ... of *terpenes*, and to restrict the name *terpene* to the compounds $C_{10}H_{16}$.

The thermal decomposition of almost all terpenes gives isoprene as one of the products, and this led to the suggestion that the skeleton structures of all naturally occurring terpenes can be built ...

by the structure of any plant-synthesised terpene. Furthermore, Ingold (1925) pointed out that the isoprene units in natural terpenes were joined "head to tail" (the head being the branched end of isoprene). This divisibility into isoprene units, and their head to tail union, may conveniently be referred to as the *special isoprene rule*.

It should be noted, however, that this rule, which has proved very useful, can only be used as a guiding principle and not as a fixed rule. Several exceptions to it occur among the simpler terpenes, *e.g.*, lavandulol is composed of two isoprene units which are not joined head to tail; also, the carotenoids are joined tail to tail at their centre (see Ch. IX).

```
CH3\                          /CH2
     C=CH·CH2·CH—C                              CH3
CH3/           |   \CH3                          |
             CH2OH                     CH2=C—CH=CH2
        lavandulol                        isoprene
```

The carbon skeletons of open-chain monoterpenes and sesquiterpenes are:

```
     C          :     C
     |          :     |
C—C—C—C—C  +  C—C—C—C
head        tail  :  head        tail
                   :
     C          :     C                  C
     |          :     |          :        |
C—C—C—C—C  +  C—C—C—C—C  +  C—C—C—C
                   :                  :
```

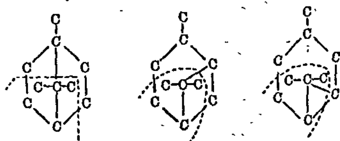
Monocyclic terpenes contain a six-membered ring, and in this connection Ingold (1921) pointed out that a *gem*-dialkyl group tends to render the *cyclo*hexane ring unstable. Hence, in closing the open chain to a *cyclo*hexane ring, use of this "*gem*-dialkyl rule" limits the number of possible structures (but see, *e.g.*, abietic acid, §31). Thus the monoterpene open chain can give rise to only *one* possibility for a monocyclic monoterpene, *viz.*, the *p*-cymene structure. This is shown in the following structures, the acyclic structure being written in the conventional "ring shape".

```
        C                         C
        |                         |
        C                         C
      /   \                     /   \
     C     C                   C     C
  --+--    |                 --+.    |
     C     C                   C  ` .C
      \                         \  / `.
        C                         C
        |                         |
        C                         C
      /   \                     /   \
     C     C                   C     C

   acyclic                    p-cymene
   structure                  structure
```

All natural monocyclic monoterpenes are derivatives of *p*-cymene.

Bicyclic monoterpenes contain a six-membered ring and a three-, four-, or five-membered ring. Ingold (1921) also pointed out that *cyclo*propane and *cyclo*butane rings require the introduction of a *gem*-dimethyl group to render them sufficiently stable to be capable of occurrence in nature. Thus closure of the C_{10} open chain gives three possible bicyclic structures; all three types are known.



If we use these ideas with the sesquiterpene acyclic structure, then we find that only three monocyclic and three bicyclic structures are possible (not all are known; see the sesquiterpenes).

§2. Isolation of monoterpenes and sesquiterpenes. Plants containing essential oils usually have the greatest concentration at some particular time, *e.g.*, jasmine at sunset. In general, there are

four methods of extraction of the terpenes: (i) expression; (ii) steam distillation; (iii) extraction by means of volatile solvents; (iv) adsorption in purified fats (*enfleurage*). Method (ii) is the one most widely used; the plant is macerated and then steam distilled. If the compound decomposes under these conditions, it may be extracted with light petrol at 50°, and the solvent then removed by distillation under reduced pressure. Alternatively, the method of adsorption in fats is used. The fat is warmed to about 50°, and then the flower petals are spread on the surface of the fat until the latter is saturated. The fat is now digested with ethanol, any fat that dissolves being removed by cooling to 20°. The essential oils so obtained usually contain a number of terpenes, and these are separated by fractional distillation. The terpene hydrocarbons distil first, and these are followed by the oxygenated derivatives. Distillation of the residue under reduced pressure gives the sesquiterpenes, and these are separated by fractional distillation.

§3. General methods of determining structure. The following brief account gives an indication of the various methods used in elucidating the structures of the terpenes (see the text for details).

(i) A pure specimen is obtained, and the molecular formula is ascertained by the usual methods. If the terpene is optically active, its specific rotation is measured. Optical activity may be used as a means of distinguishing structures (see, *e.g.*, §12).

(ii) If oxygen is present in the molecule, its functional nature is ascertained, *i.e.*, whether it is present as hydroxyl, aldehyde, ketone, etc. (*cf.* alkaloids, §4. XIV).

(iii) The presence of olefinic bonds is ascertained by means of bromine, and the number of double bonds is determined by analysis of the bromide, or by quantitative hydrogenation, or by titration with monoperphthalic acid. These facts lead to the molecular formula of the parent hydrocarbon, from which the number of rings present in the structure may be deduced.

(iv) The preparation of nitrosochlorides and a study of their behaviour (see also the nitroso compounds, Vol. I).

(v) Dehydrogenation of terpenes with sulphur or selenium, and an examination of the products thereby obtained (see also §2 vii. X).

(vi) Measurement of the refractive index leads to a value for the molecular refractivity. From this may be deduced the nature of the carbon skeleton (see, in particular, sesquiterpenes). Also, optical exaltation indicates the presence of double bonds in conjugation (*cf.* §11. I).

(vii) Measurement of the ultraviolet and infra-red spectra.

(viii) Degradative oxidation. The usual reagents used for this

purpose are ozone, acid or alkaline permanganate, chromic acid, and sodium hypobromite. In general, degradative oxidation is the most powerful tool for elucidating the structures of the terpenes.

(ix) After the analytical evidence has led to a tentative structure (or structures), the final proof of structure depends on synthesis. In terpene chemistry, many of the syntheses are ambiguous, and in such cases analytical evidence is used *in conjunction* with the synthesis. Many terpenes have not yet been synthesised.

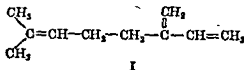
MONOTERPENES

The monoterpenes may be subdivided into three groups: acyclic, monocyclic and bicyclic. This classification affords a convenient means of study of the monoterpenes.

ACYCLIC MONOTERPENES

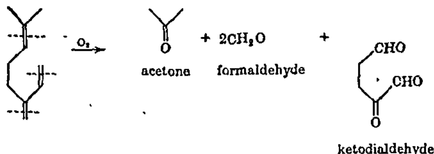
§4. **Myrcene,** $C_{10}H_{16}$, is an acyclic monoterpene hydrocarbon which occurs in verbena and bay oils. It is a liquid, b.p. 166–168°. Catalytic hydrogenation (platinum) converts myrcene into a decane, $C_{10}H_{22}$; thus myrcene contains three double bonds, and is an open-chain compound. Furthermore, since myrcene forms an adduct with maleic anhydride, two of the double bonds are conjugated (Diels *et al.*, 1929; see the Diels–Alder reaction, Vol. I). This conjugation is supported by evidence obtained from the ultraviolet spectrum of myrcene (Booker *et al.*, 1940). These facts, *i.e.*, that myrcene contains three double bonds, two of which are in [illegible] These results can be explained by assigning stru[illegible]

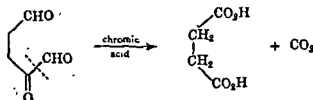
I to myrcene. In terpene chemistry it has become customary to use conventional formulae rather than those of the type I. In these conventional formulae only lines are used; carbon atoms are at the junctions of pairs of lines or at the end of a line, and unsaturation is indicated by double bonds. Furthermore, the carbon skeleton is usually drawn in a ring fashion (the *cyclo*hexane ring).


I


II

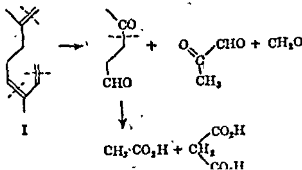
Thus myrcene may be represented as II, and this type of structural formula will, in general, be used in this book. Thus the process of ozonolysis and oxidation of the ketodialdehyde may be written:

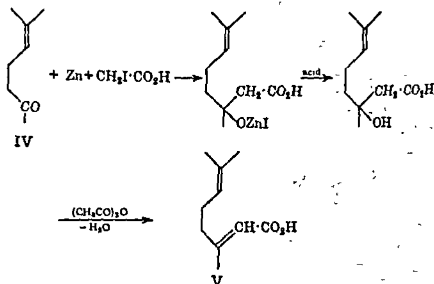




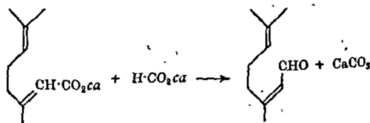
This structure for myrcene is supported by the fact that on hydration (under the influence of sulphuric acid), myrcene forms an alcohol which, on oxidation, gives citral. The structure of this compound is known (see §5), and its formation is in accord with the structure given to myrcene.

§4a. Ocimene, $C_{10}H_{16}$, b.p. 81°/30 mm. When catalytically hydrogenated, ocimene adds on three molecules of hydrogen to form a decane. Thus ocimene is an acyclic compound which contains three double bonds. Furthermore, since ocimene forms an adduct with maleic anhydride, two of the double bonds are conjugated. On ozonolysis, ocimene produces formaldehyde, methylglyoxal, lævulaldehyde, acetic and malonic acids, and some acetone. All of these products, except acetone, are accounted for by structure I

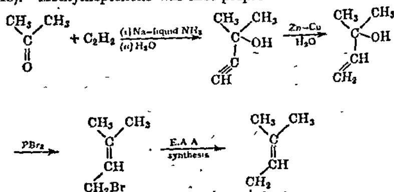




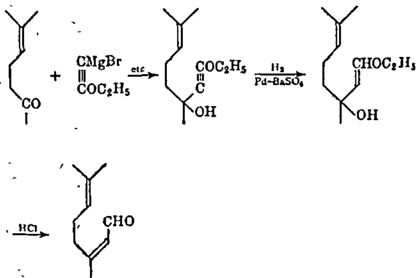
(1898) by distilling a mixture of the calcium salts of geranic and formic acids (*ca* represents "half an atom of calcium"):



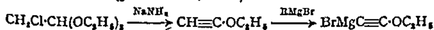
A more recent synthesis of citral is that of Arens and van Dorp (1948). Methylheptenone was first prepared as follows:



Then the methylheptenone was treated with ethoxyacetylenemagnesium bromide, the product reduced and then de-alkylated.

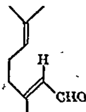


It should be noted that an *allylic rearrangement* occurs in both parts of this synthesis (see also §8). Ethoxyacetylenemagnesium bromide may conveniently be prepared from chloroacetaldehyde diethyl acetal as follows (Jones *et al.*, 1954):

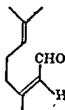


$$CH_2Cl{\cdot}CH(OC_2H_5)_2 \xrightarrow{NaNH_2} CH{\equiv}C{\cdot}OC_2H_5 \xrightarrow{RMgBr} BrMgC{\equiv}C{\cdot}OC_2H_5$$

Examination of the formula of citral shows that two geometrical isomers are possible:



trans-form;
citral-*a*;
geranial

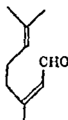


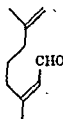
cis-form;
citral-*b*;
neral

Both isomers occur in natural citral, *e.g.*, *two* semicarbazones are formed by citral; both forms of citral itself have also been obtained: **citral-a** (also known as *geranial*) has a b.p. 118–119°/20 mm., and **citral-b** (also known as *neral*) has a b.p. 117–118°/20 mm. The configurations of these two forms have been determined from

a consideration of the ring closures of the corresponding alcohols (see *geraniol*, §7).

The problem of the structure of citral is further complicated for the following reasons. Ozonolysis of citral gives acetone, lævulaldehyde and glyoxal (Harries, 1903, 1907); these products are to be expected from structure III. On the other hand, Grignard *et al.* (1924) also isolated a small amount of formaldehyde from the products of ozonolysis; this points towards structure VI, which has an *iso*propenyl end-group. Thus citral has been regarded as



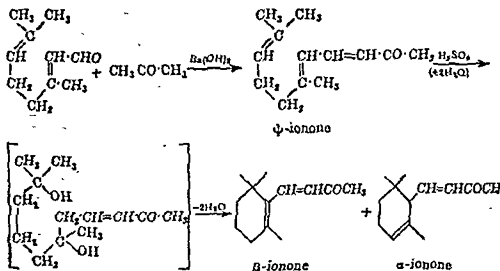


a mixture of *four* substances, two geranials and two nerals. Assuming, then, that both the *iso*propylidene and *iso*propenyl forms are present, it is possible that these two structures form a three-carbon tautomeric system:



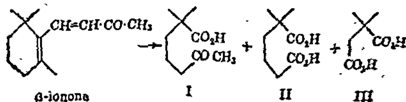
Recent work, however, has cast doubt on the existence of these two forms in citral. According to infra-red spectroscopic studies, it appears that naturally occurring acyclic monoterpenes *as a class* possess only the *iso*propylidene end-group structure (Barnard, Bateman *et al*, 1950). According to these authors, *during oxidative degradation*, partial rearrangement from the *iso*propylidene to the *iso*propenyl structure occurs, and so this method of determining fine structure is unreliable (see also *geraniol*, §7).

§6. **Ionones.** When citral is condensed with acetone in the presence of barium hydroxide, ψ-ionone is formed and this, on heating with dilute sulphuric acid in the presence of glycerol, forms a mixture of α- and β-ionones (Tiemann and Kruger, 1893). The proportion of α to β varies with the nature of the cyclising agent used, *e.g.*, with sulphuric acid, β-ionone is the main product; with phosphoric acid, α-ionone is the main product. Both ionones have been obtained from natural sources; the β-isomer is optically inactive, whereas the α-isomer can exist in optically active forms

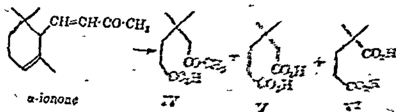


since it contains one asymmetric carbon atom. *Actually, the* (+)-, (−)-, and (±)-forms of α-ionone occur naturally. Very dilute ethanolic solutions of β-ionone have the odour of violets.

The structures of the ionones were established by a study of the oxidation products produced by potassium permanganate (Tiemann, 1898, 1900); β-ionone gave geronic acid, I, α:α-dimethyladipic

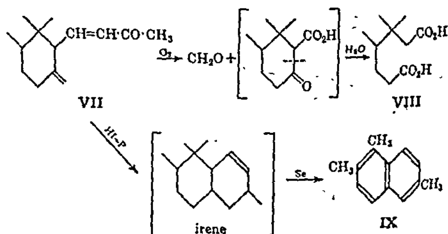


acid, II, and α:α-dimethylsuccinic acid, III. On the other hand, α-ionone gave a mixture of *isogeronic* acid, IV, β:β-dimethyladipic acid, V, and α:α-dimethylglutaric acid, VI.

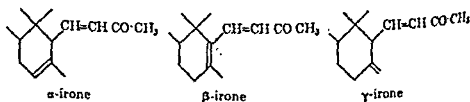


The ionones are [illegible], $C_{13}H_{20}O$; [illegible] oil obtained from [illegible]. The [illegible]

established by Ruzicka *et al.* (1947), who showed that on ozonolysis, irone gives formaldehyde and β:β:γ-trimethylpimelic acid, VIII; also, reduction of irone with hydriodic acid and red phosphorus, followed by dehydrogenation with selenium, gives 1.2:6-trimethylnaphthalene, IX. Ruzicka therefore proposed structure VII for



irone. Ruzicka (1947) further showed that irone was a mixture of three isomers (VII is γ-irone):

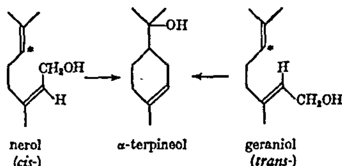


α-irone β-irone γ-irone

§7. **Geraniol,** $C_{10}H_{18}O$, b.p. 229–230°/757 mm. This is found in many essential oils, particularly rose oil. Geraniol was shown to be a primary alcohol, *e.g.*, on oxidation it gives an aldehyde (citral-a); and since it forms a tetrabromide, geraniol therefore contains two double bonds. Reduction of citral produces geraniol, but at the same time some nerol is formed. The *structural*

pt the double bonds to be 2 and 7 (*cf.* citral, §5). Thus geraniol and nerol are geometrical isomers. Geraniol has been assigned the

trans configuration and nerol the *cis* on the fact that cyclisation to α-terpineol (§11) by means of dilute sulphuric acid takes place about 9 times as fast with nerol as it does with geraniol; this faster rate with nerol is due to the proximity of the alcoholic group to the carbon (*) which is involved in the ring formation. Thus:

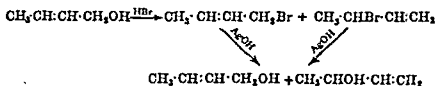


nerol (*cis*-) α-terpineol geraniol (*trans*-)

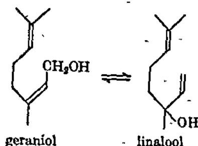
Nerol also occurs naturally in various essential oils, *e.g.*, oil of neroli, bergamot, etc.; its b.p. is 225–226°.

Knights *et al.* (1955) have found that, on ozonolysis, geranyl acetate gives less than 3 per cent. of formaldehyde, and have concluded that the acetate and geraniol itself have predominantly the *iso*propylidene structure (*cf.* citral, §5).

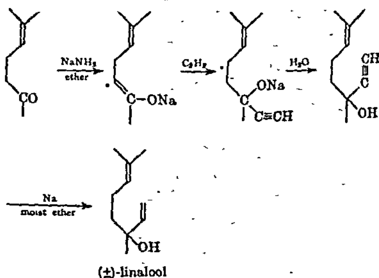
§8. **Linalool,** $C_{10}H_{18}O$, b.p. 198–199°. This is an optically active compound; the (−)-form occurs in rose oil and the (+)-form in orange oil. It was shown to be a tertiary alcohol, and since it adds on two molecules of hydrogen on catalytic hydrogenation, it must contain two double bonds. When heated with acetic anhydride, linalool is converted into geranyl acetate; and the latter is converted into the former by heating with steam at 200° under pressure. Also, heating linalool with hydrogen chloride in toluene solution at 100° produces geranyl chloride, and this, when treated with moist silver oxide in benzene solution, is reconverted into linalool. These reactions are parallel to those which occur when crotyl alcohol is treated with hydrogen bromide; a mixture of crotyl bromide and methylvinylcarbinyl bromide is obtained. When *either* of these products is treated with moist silver oxide, a mixture of crotyl alcohol and methylvinylcarbinol is obtained.



Thus the elucidation of the structure of linalool is complicated by the ease with which the *allylic rearrangement* occurs (see also Vol. I). Since the structure of geraniol is known, a possible structure for linalool is obtained on the basis of this allylic rearrangement.



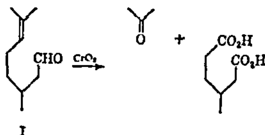
This structure has been confirmed by synthesis of linalool (Ruzicka *et al.*, 1919), 6-methylhept-5-en-2-one was treated as follows:



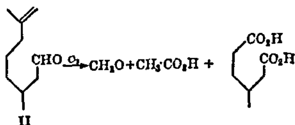
§9. **Citronellal,** $C_{10}H_{18}O$. This is an optically active compound which occurs in citronella oil. *Citronellal is an aldehyde*; reduction with sodium amalgam converts it into the alcohol citronellol, $C_{10}H_{20}O$, and oxidation gives citronellic acid, $C_{10}H_{18}O_2$. Now there is another aldehyde, **rhodinal**, which is isomeric with citronellal, and on reduction, rhodinal gives the alcohol, rhodinol, which is isomeric with citronellol. Furthermore, reduction of ethyl geranate with sodium and ethanol gives rhodinol (Bouveault *et al.*, 1900).

Oxidation of citronellal with chromic acid gives β-methyladipic acid and acetone (Tiemann *et al.*, 1896, 1897). Rhodinal also gives

the *same* products on oxidation. Thus structure I would fit the facts for both citronellal and rhodinal. On the other hand, ozonolysis of citronellal gives β-methyladipic acid, acetone and some

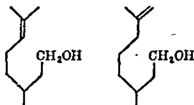


formaldehyde (Harries *et al.*, 1908). These results point towards structure II for citronellal, as well as I. Thus citronellal appears to be a mixture of I (*iso*propylidene end-group) and II (*iso*propenyl end-group). Furthermore, a detailed study of rhodinal has shown



that this compound is identical with citronellal, but consists of a mixture of the two forms in different proportions (but *cf.* citral, §5).

§9a. Citronellol and Rhodinol, $C_{10}H_{20}O$. (−)-Citronellol occurs in rose and geranium oils, and is a mixture of the two forms:

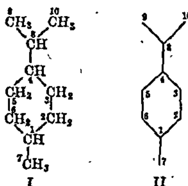


The (+)-form of citronellol is made commercially by reduction of citronellal with sodium or aluminium amalgam; it also occurs in Java citronella oil.

Rhodinol is identical with citronellol, but the proportions of the two forms are different from those which occur in citronellol; the identity of citronellol and rhodinol is shown by the products of ozonolysis.

MONOCYCLIC MONOTERPENES

§10. Nomenclature. For the purposes of nomenclature of the monocyclic monoterpenes, the fully saturated compound *p*-methyl-*iso*propyl*cyclo*hexane, hexahydro-*p*-cymene, or *p*-menthane, $C_{10}H_{20}$, is used as the parent substance; it is a synthetic compound, b p. 170°. *p*-Menthane is I, and II is a conventional method of drawing formula I. The positions of substituents and double bonds are



indicated by numbers, the method of numbering being shown in I (and II). When a compound derived from *p*-menthane contains one or more double bonds, ambiguity may arise as to the position of a double bond when this is indicated in the usual way by a number which locates the *first* carbon atom joined by the double bond. To prevent ambiguity, the *second* carbon atom joined to the double bond is also shown, but is placed in parentheses. The following examples illustrate the method of nomenclature; in the first example, all the types of methods of nomenclature have been given; in the second and third examples, only the nomenclature that will be used in this book is given.



Δ^2-*p*-menthene; 2-*p*-menthene; *p*-menth-2-ene; *p*-menthene-2.



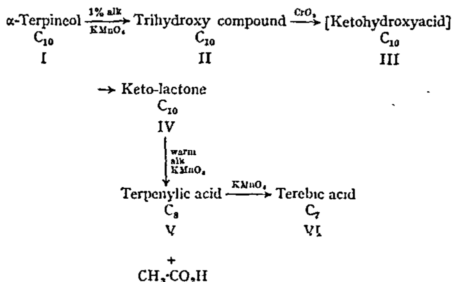
p-menth-1(7)-ene



p-mentha-1:4(8)-diene

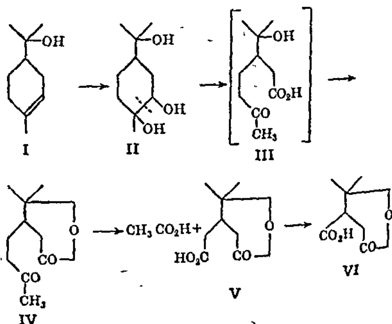
§11. α-Terpineol. This is an optically active monoterpene that occurs naturally in the (+)-, (−)- and (±)-forms; it is a solid, m.p. (of the racemic modification) 35°. The molecular formula of α-terpineol is $C_{10}H_{18}O$, and the oxygen atom is present as a tertiary alcoholic group (as shown by the reactions of α-terpineol). Since α-terpineol adds on two bromine atoms, it therefore contains one double bond. Thus the parent (saturated) hydrocarbon of α-terpineol has the molecular formula $C_{10}H_{20}$. This corresponds to C_nH_{2n}, the general formula of the (monocyclic) *cyclo*alkanes, and so it follows that α-terpineol is a monocyclic compound.

When heated with sulphuric acid, α-terpineol forms some *p*-cymene. Taking this in conjunction with the tentative proposal that α-terpineol is monocyclic, it is reasonable to infer that α-terpineol contains the *p*-cymene skeleton. Thus we may conclude that α-terpineol is probably *p*-menthane with one double bond and a tertiary alcoholic group. The positions of these functional groups were ascertained by Wallach (1893, 1895) by means of *graded* oxidation. The following chart gives the results of Wallach's work; only the carbon content is indicated to show the fate of these carbon atoms (the formulae are given in the text).



Oxidation of α-terpineol, I, with 1 per cent. alkaline potassium permanganate hydroxylates the double bond to produce the trihydroxy compound II, $C_{10}H_{20}O_3$. This, on oxidation with chromic acid (chromium trioxide in acetic acid), produces a compound with

the molecular formula $C_{10}H_{16}O_3$ (IV). This compound was shown to contain a ketonic group, and that it was neutral, *e.g.*, it gave no reaction with sodium carbonate solution. When, however, IV was refluxed with excess of standard sodium hydroxide solution, and then back titrated, it was found that alkali had been consumed, the amount corresponding to the presence of one carboxyl group. Thus compound IV appears to be the *lactone* of a monocarboxylic acid. Furthermore, since it is the lactone that is isolated and not the hydroxy acid, this *spontaneous* lactonisation may be interpreted as being produced from a γ-hydroxyacid, *i.e.*, IV is a γ-lactone, and therefore III is a γ-hydroxyacid. It is possible, however, for δ-hydroxyacids to spontaneously lactonise, and so whether IV is a γ- or δ-lactone is uncertain at this stage of the evidence. Now, since IV is formed from II by scission of the glycol bond, and since

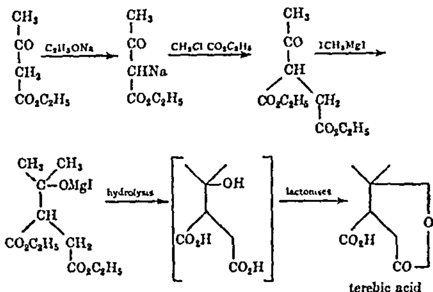


there is *no loss of carbon atoms* in the process, the double bond must therefore be in the ring in I. On warming with alkaline permanganate, IV gave acetic acid and a compound $C_8H_{12}O_4$ (V). The formation of acetic acid suggests that IV is a *methyl ketone*, *i.e.*, a $CH_3{\cdot}CO$ group is present. Thus IV is a methyl ketone and a lactone; it is known as homoterpenyl methyl ketone, and the structure assigned to it has been confirmed by synthesis (Simonsen *et al.*, 1932). A study of the properties of terpenylic acid, V, showed that it was the lactone of a monohydroxydicarboxylic acid. Further

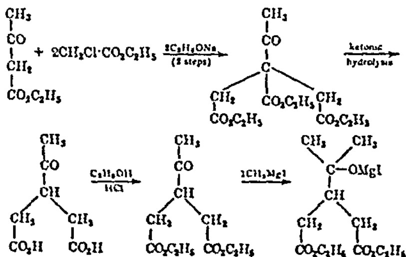
oxidation of terpenylic acid gives terebic acid $C_7H_{10}O_4$ (VI), which is also the lactone of a monohydroxydicarboxylic acid.

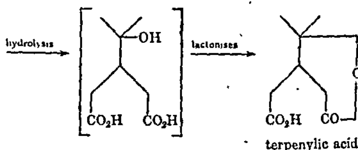
The above reactions can be formulated as shown, *assuming* I (*p*-menth-1-en-8-ol) as the structure of α-terpineol. These reactions were formulated by Wallach, who adopted formula I which had been proposed by Wagner (1894). The structures of terpenylic (V) and terebic (VI) acids were established by synthesis, *e.g.*, those of Simonsen (1907).

Terebic acid, m.p. 175°.



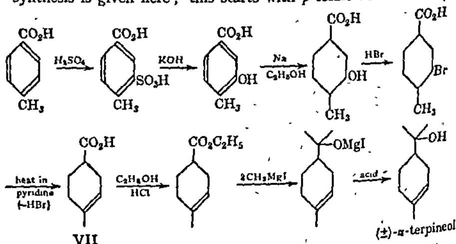
Terpenylic acid, m.p. 90°.



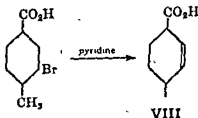


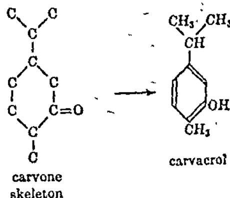
It is of interest to note here that Sandberg (1957) has prepared the β-acetotricarballylate in *one* step from acetoacetic ester and ethyl bromoacetate in the presence of sodium hydride (in benzene solution).

These syntheses strengthen the evidence for the structure assigned to α-terpineol, but final proof rests with a synthesis of α-terpineol itself. This has been carried out by Perkin, junior (1904), and by Perkin, junior, with Meldrum and Fisher (1908). Only the second synthesis is given here; this starts with p-toluic acid.

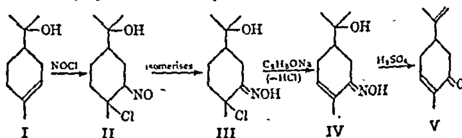


Compound VII was also resolved with strychnine, each enantiomorph treated as shown above (esterified, etc.), and thereby resulted in the formation of (+)- and (−)-terpineol. It should be noted that in the above synthesis the removal of a molecule of hydrogen





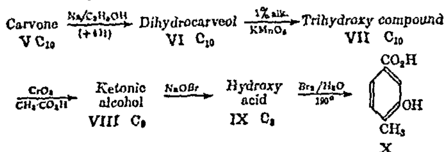
The structure of carvone is largely based on the fact that *carvone* may be prepared from α-terpineol as follows:



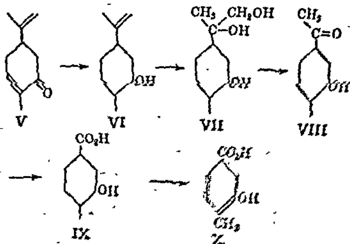
The addition of nitrosyl chloride to α-terpineol, I, produces α-terpineol nitrosochloride, II, the addition occurring according to Markownikoff's rule (the chlorine is the negative part of the addendum; see Vol. I). This nitrosochloride rearranges spontaneously to the oximino compound, III (see nitroso-compounds, Vol. I; it might be noted that this rearrangement proves the orientation of the addition of the nitrosyl chloride to the double bond; addition the other way could not give an oxime, since there is no hydrogen atom at position 1 in α-terpineol). Removal of a molecule of hydrogen chloride from III by means of sodium ethoxide produces IV, and this, on warming with dilute sulphuric acid, loses a molecule of water with simultaneous hydrolysis of the oxime to form carvone, V. Thus, according to this interpretation of the reactions, carvone is *p*-menth-6:8-dien-2-one. Actually, these reactions show that carvone has the same carbon skeleton as α-terpineol, and also confirm the position of the keto group. They do not prove conclusively the positions of the two double bonds; instead of position 6 (in IV), the double bond could have been 1(7), and instead of position 8 (as in V), the double bond could have been 4(5). Thus the above reactions constitute an ambiguous synthesis of carvone (α-terpineol has already been synthesised). The exact positions

of these two double bonds have been determined *analytically* as follows.

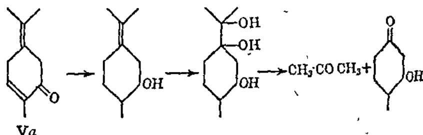
The double bond in the 8-position. The following reactions were carried out by Tiemann and Semmler (1895).



Reduction of carvone, V, with sodium and ethanol gives dihydrocarveol, $C_{10}H_{18}O$ (*VI*); *this is a secondary alcohol and contains one* double bond, *i.e.*, the keto group and *one* of the two double bonds in carvone have been reduced. Hydroxylation of the double bond in dihydrocarveol by means of 1 per cent. alkaline permanganate produces the trihydroxy compound $C_{10}H_{20}O_3$ (VII). Oxidation of VII with chromic acid causes scission of the glycol bond to produce a compound $C_9H_{16}O_2$ (*VIII*); this was shown to contain a keto group and a hydroxyl (alcoholic) group. The action of sodium hypobromite on VIII caused the loss of one carbon atom to produce the compound $C_8H_{14}O_3$ (IX); this was shown to be a hydroxymonocarboxylic acid, and since *one* carbon is lost in its formation, its precursor VIII must therefore be a methyl ketone. Finally, dehydrogenation of IX by heating with bromine-water at 190° under pressure produced *m*-hydroxy-*p*-toluic acid, X (a *known* compound). Tiemann and Semmler explained these reactions on the assumption that one double bond in carvone is in the 8-position. Thus:

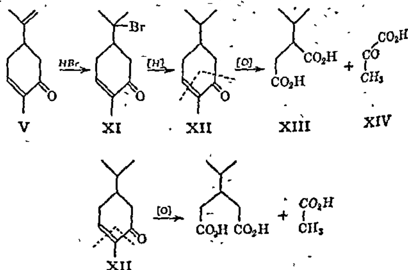


Had the double bond been in the 4(8)-position (structure V*a*), then compound VIII, and consequently X, could not have been obtained, since *three* carbon atoms would have been lost during the oxidation.



It might be noted in passing that V contains an asymmetric carbon atom, whereas V*a* is a symmetrical molecule and so cannot exhibit optical activity. Since carvone is known in optically active forms, structure V*a* must be rejected on these grounds.

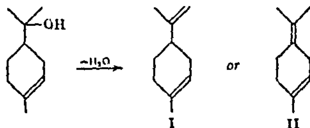
The double bond in the 6-position. Carvone adds on one molecule of hydrogen bromide to form carvone hydrobromide, $C_{10}H_{15}OBr$ (XI), and this, on treatment with zinc dust and methanol, is converted into carvotanacetone, $C_{10}H_{16}O$ (XII), by replacement of the bromine atom by hydrogen. Thus the final result of these reactions is to saturate *one* of the two double bonds in carvone. Carvotanacetone, on oxidation with permanganate, gives *iso*propylsuccinic acid, XIII, and pyruvic acid, XIV (Semmler, 1900). These products are obtainable only if the ring contains the double bond in



the 6-position. Had the double bond been in the 1(7)-position, formic acid and not pyruvic acid would have been obtained. Further support for the 6-position is provided by the work of Simonsen *et al.* (1922), who obtained *β-iso*propylglutaric acid and acetic acid on oxidation of carvotanacetone with permanganate.

§13. Limonene, $C_{10}H_{16}$, b.p. 175·5–176·5°. This is optically active; the (+)-form occurs in lemon and orange oils, the (−)-form in peppermint oil, and the (±)-form in turpentine oil. The racemic modification is also produced by racemisation of the optically active forms at about 250°. The racemic modification is also known as **dipentene**; this name was given to the inactive form before its relation to the active form (limonene) was known.

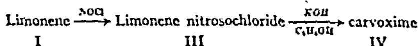
Since limonene adds on four bromine atoms, it therefore contains two double bonds. (+)-Limonene may be prepared by dehydrating (+)-α-terpineol with potassium hydrogen sulphate, and limonene (or dipentene) may be converted into α-terpineol on shaking with dilute sulphuric acid. Thus the carbon skeleton and the position



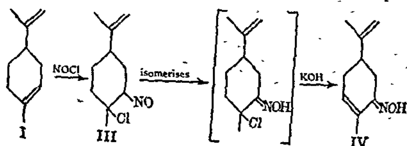
of one double bond in limonene are known. The position of the other double bond, however, remains uncertain from this preparation; I or II is possible.

Proof for position 8. Structure I contains an asymmetric carbon atom (C_4), and hence can exhibit optical activity. II is a symmetrical molecule and so cannot be optically active. Therefore I must be limonene.

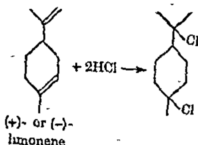
Chemical proof for position 8 is afforded by the following reactions:



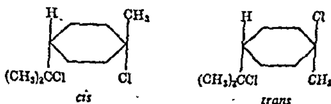
Since the structure of carvoxime is known, it therefore follows that I must have one double bond in position 8; thus the above reactions may be written:



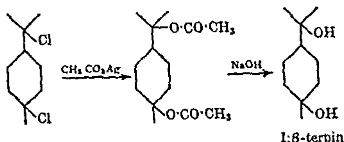
The connection between limonene and dipentene is shown by the fact that (+)- or (−)-limonene adds on two molecules of hydrogen chloride in the presence of moisture to form limonene dihydrochloride, and this is identical with dipentene dihydrochloride.



Limonene dihydrochloride no longer contains an asymmetric carbon atom, and so is optically inactive. It can, however, exhibit geometrical isomerism; the *cis*-form is produced from limonene, and the *trans*-form from cineole (§14).

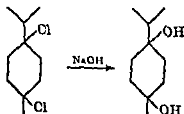


Dipentene can be regenerated by heating the dihydrochloride with sodium acetate in acetic acid, or boiling with aniline. On the other hand, when limonene dihydrochloride is heated with silver acetate in acetic acid, and then hydrolysing the ester with sodium hydroxide, 1:8-terpin is formed; the direct action of sodium hydroxide on the dihydrochloride regenerates dipentene.



1:8-Terpin exists in two geometrical isomeric forms, corresponding to the *cis* and *trans* dipentene dihydrochlorides. *cis*-1:8-Terpin is the common form, m p. 105°, and readily combines with *one* molecule of water to form terpin hydrate. The *trans* form, m p. 158–159°, does not form a hydrate.

There is also a **1:4-terpin**; this was originally prepared by the action of dilute alkali on terpinene dihydrochloride.



Terpinenes, $C_{10}H_{16}$. There are three isomeric terpinenes, and all give the same terpinene dihydrochloride with hydrogen chloride.

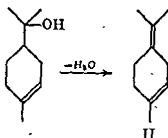

α-terpinene
b.p. 180-182°


β-terpinene
b.p. 173-174°


γ-terpinene
b.p. 69-73°/20mm.

All three occur naturally.

Terpinolene, $C_{10}H_{16}$, b.p. 67–68°/10 mm. This occurs naturally. It is not optically active, and since it may be prepared by dehydrating α-terpineol with oxalic acid, its structure is known (it is II, the alternative formula offered for limonene). Terpinolene adds on two molecules of hydrogen chloride to form dipentene dihydrochloride.



Phellandrenes, $C_{10}H_{16}$. There are two phellandrenes, both of which are optically active, and all the enantiomorphs occur naturally.

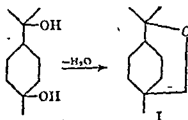


α-phellandrene
b.p. 58-59°/16 mm.



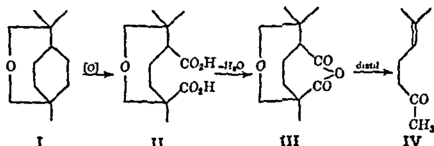
β-phellandrene
b.p. 171-172°

§14. **1:8-Cineole**, $C_{10}H_{18}O$, *b.p.* 174·4°. This occurs in eucalyptus oils. It is isomeric with α-terpineol, but contains neither a hydroxyl group nor a double bond. The oxygen atom in cineole is inert, *e g.*, it is not attacked by sodium or by the usual reducing agents. This inertness suggests that the oxygen atom is of the ether type. Support for this is obtained from the fact that dehydration of *cis*-1:8-terpin gives 1:8-cineole; at the same time, this reaction suggests that the structure of cineole is I.



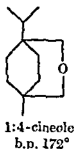
Further support for this structure is afforded by a study of the products obtained by oxidation (Wallach *et al.*, 1888, 1890, 1892). When oxidised with potassium permanganate, cineole forms cineolic acid, II, and this, on distillation with acetic anhydride, forms cineolic anhydride, III. When distilled at atmospheric pressure,

cineolic anhydride forms 6-methylhept-5-en-2-one, IV, a *known* compound (§5). These reactions were interpreted by Wallach as follows:



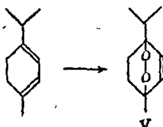
Further work on the structure of cineolic acid has confirmed the above sequence of reactions (Rupe, 1901, —).

There is also a 1:4-cineole; this occurs naturally.



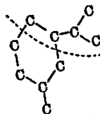
1:4-cineole
b.p. 172°

Ascaridole, $C_{10}H_{16}O_2$, b.p. 96–97°/8 mm. The cineoles are oxides; ascaridole, however, is a peroxide, the only known terpene peroxide, and it occurs naturally in, *e.g.*, chenopodium oil. When heated to 130–150°, ascaridole decomposes with explosive violence. When reduced catalytically, ascaridole forms 1:4-terpin (Wallach,



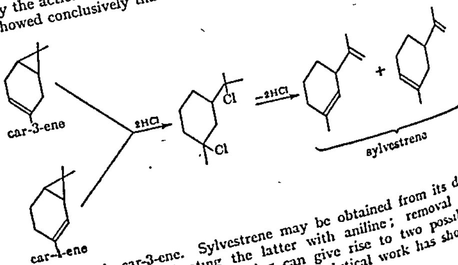
1912), and this led to the suggestion that ascaridole is V. This structure has been confirmed by further analytical work. Ascaridole has been synthesised by Ziegler *et al.* (1944) by the irradiation of α-terpinene in dilute solution in the presence of chlorophyll.

§15. Sylvestrene, $C_{10}H_{16}$, b.p. 175–178°. This compound exists in (+)-, (−)-, and (±)-forms; the racemic modification is also known as **carvestrene** (*cf.* limonene and dipentene, §13). The (+)-form of sylvestrene was first obtained from Swedish pine needle oil (Attenberg, 1877), and was shown to contain the *m*-cymene carbon skeleton (Baeyer *et al.*, 1898). Thus sylvestrene appeared to be the only monocyclic monoterpene which did not have the *p*-cymene structure and was obtainable from natural sources. Although the *m*-cymene structure can be divided into two isoprene units (Wallach's isoprene rule), these two units are not joined head to tail.



m-cymene skeleton

Subsequent work, however, showed that sylvestrene does not occur in pine oil. In the extraction of sylvestrene, the pine oil is heated with hydrogen chloride to give sylvestrene dihydrochloride. This compound was shown by Simonsen *et al.* (1923, 1925) to be produced by the action of hydrogen chloride on car-3-ene, *i.e.*, these workers showed conclusively that the terpene originally present in Swedish

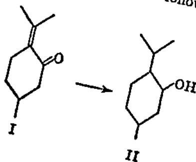


pine oil is car-3-ene. Sylvestrene may be obtained from its dihydrochloride by heating the latter with aniline; removal of hydrogen chloride from the ring can give rise to two possible positions for the ring double bond. Analytical work has shown

that the side-chain is *iso*propenyl (and not *iso*propylidene), and that sylvestrene is a mixture of the two forms, *m*-mentha-1:8-diene and *m*-mentha-6.8-diene. Furthermore, it has been shown that car-4-ene is also present in pine oil; both of these carenes are readily converted into sylvestrene, and so it appears that the precursor of sylvestrene (itself a mixture) is a mixture of the two carenes (see §21).

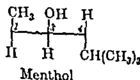
The enantiomorphs of sylvestrene have been synthesised (Perkin, junior, *et al.*, 1913), and it has also been shown that an equimolecular mixture of the dihydrochlorides of (+)- and (−)-sylvestrene is identical with carvestrene dihydrochloride.

§16. Menthol and menthone. Menthol, $C_{10}H_{20}O$, is an optically active compound, but only the (−)-form occurs naturally, *e.g.*, in peppermint oils. (−)-Menthol, m.p. 34°, is a saturated compound, and the functional nature of the oxygen atom is alcoholic, as shown by its reactions, *e.g.*, menthol forms esters. Furthermore, since oxidation converts menthol into menthone, a *ketone*, the alcoholic group in menthol is therefore secondary. Also, since reduction with hydrogen iodide gives *p*-menthane, menthol most probably contains this carbon skeleton. Finally, since (+)-pulegone gives menthol on reduction, and since the structure of pulegone is known to be I (see §17), it therefore follows that menthol must

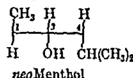


be II. This structure, *p*-menth-3-ol, for menthol has been confirmed by consideration of the oxidation products of menthone (see below), and also by the synthesis of menthol.

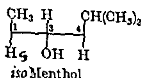
Examination of the menthol structure shows that three dissimilar asymmetric carbon atoms (1, 3 and 4) are present; thus eight optically active forms (four racemic modifications) are possible theoretically. All eight enantiomorphs are known and their configurations are as follows (the horizontal lines represent the plane of the *cyclo*hexane ring):



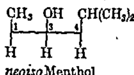
Menthol



*neo*Menthol



*iso*Menthol

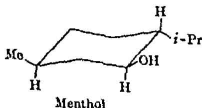


*neoiso*Menthol

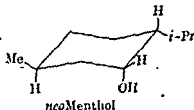
These configurations have been assigned from a study of chemical and optical relationships and the Auwers-Skita rule. More recently the application of conformational analysis has confirmed these results. Eliel (1953) applied the principle that the esterification of an axial hydroxyl group occurs less readily than with an equatorial one. Furthermore, Eliel postulated that the reaction proceeds *via* the conformation of the molecule in which the reactive hydroxyl group is equatorial, and that the rate differences should be attributed to that energy necessary to place the other substituents, if necessary, into the axial conformation (see also §12. IV). On this basis, the rates of esterification of the isomeric menthols will be:


menthol > *iso-* > *neoiso-* > *neo-*.

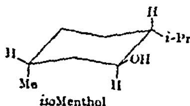
These are the orders of rates actually obtained by Read *et al.* (1934). The following conformations have been assigned by Eliel from chemical studies, and are supported by Cole *et al.* (1956) from their infra red spectra and conformation studies.



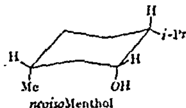
Menthol



*neo*Menthol



*iso*Menthol

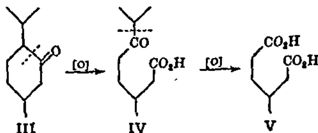


*neoiso*Menthol

In menthol, *all* of the substituents are equatorial, and in the rest

one is axial. It should also be noted that the larger of the two alkyl groups (*iso*propyl) is always equatorial (*cf.* §11. IV).

Menthone, $C_{10}H_{18}O$, b.p. 204°/750 mm. (−)-Menthone occurs in peppermint oil, and it may readily be prepared by the oxidation of (−)-menthol with chromic acid. Menthone is a saturated compound which has the characteristic properties of a ketone. When heated with hydriodic acid and red phosphorus, menthone is reduced to *p*-menthane; thus this skeleton is present in menthone. Oxidation of menthone with potassium permanganate produces a compound $C_{10}H_{18}O_3$; this compound was shown to contain a keto-group and one carboxyl group, and is known as ketomenthylic acid (IV). Ketomenthylic acid itself is very readily oxidised by permanganate to *β*-methyladipic acid (V) and some other acids (Arth, 1886; Manasse *et al.*, 1894). The foregoing oxidative reactions may be formulated as follows, on the *assumption* that III is the



structure of menthone. This structure for menthone has been confirmed by synthesis, *e.g.*, Kötz and Schwarz (1907) obtained menthone by the distillation of the calcium salt of *β*′-methyl-*α*-*iso*propylpimelic acid, which was prepared as follows. 3-Methyl*cyclo*hexanone, VI, was condensed with ethyl oxalate in the presence of sodium, and the product VII then heated under reduced pressure; this gave the ethyl ester of 4-methyl*cyclo*hexan-2-one-1-carboxylic acid, VIII. VIII, on treatment with sodium ethoxide followed by *iso*propyl iodide, gave IX, and this when boiled with ethanolic sodium ethoxide and the product then acidified, gave *β*′-methyl-*α*-*iso*propylpimelic acid, X (note the acetoacetic ester fragment in VIII).

Structure III contains two dissimilar asymmetric carbon atoms (1 and 4), and so four optically active forms (and two racemic modifications) are possible. All are known, and correspond to the menthones and *iso*menthones; these are geometrical isomers, each one existing as a pair of enantiomorphs. The configurations have been assigned on physical evidence; the *cis*-isomer has the higher refractive index and density (Auwers–Skita rule; see §5 x. IV).

VI + $(CO_2C_2H_5)_2$ $\xrightarrow{Na}$ VII ($CO_2C_2H_5$, CO) $\xrightarrow[\text{(reduced pressure)}]{\text{heat}}$ VIII ($CO_2C_2H_5$) $\xrightarrow[\text{(ii) } (CH_3)_2CHI]{\text{(i) } C_2H_5ONa}$

IX ($CO_2C_2H_5$) $\xrightarrow[\text{(ii) HCl}]{\text{(i) } C_2H_5ONa}$ X (CO_2H, $CH_2{\cdot}CO_2H$) $\xrightarrow{\text{Ca salt}}$ III

$(CH_3)_2CH$, O, CH_3, H, H — *cis*-isomer *iso*Menthone

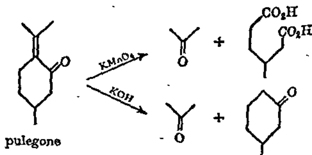
$(CH_3)_2CH$, O, H, H, CH_3 — *trans*-isomer Menthone

§17. (+)-Puleg

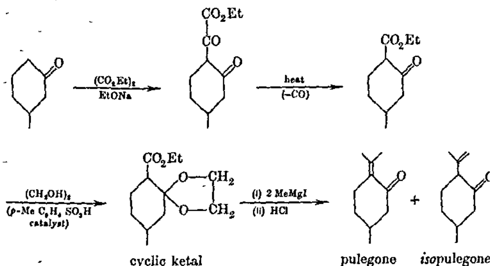
pennyroyal oils.

as a ketone. On

on further reduction, gives menthol. When oxidised with permanganate, pulegone forms acetone and β-methyladipic acid (Semmler, 1892); when boiled with aqueous ethanolic potassium hydroxide, acetone and 3-methyl*cyclo*hexanone are obtained (Wallach, 1896). These reactions show that pulegone is *p*-menth-4(8)-en-3-one.

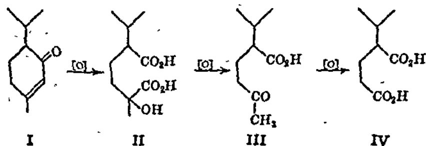


This structure has been confirmed by synthesis, starting from 3-methyl*cyclo*hexanone (Black *et al.*, 1956: *cf.* menthone, §16).



*iso*Pulegone can be isomerised to pulegone by alkaline reagents (Kon *et al.*, 1927), and Black *et al.* found that, on treating their mixture with sodium ethoxide, the resulting compound was pure pulegone.

§18. (–)-Piperitone, $C_{10}H_{16}O$, b.p. 232–233°/768 mm. This occurs in eucalyptus oils, and is a valuable source of menthone and thymol. Piperitone contains one double bond, and behaves as a ketone. Piperitone, on catalytic hydrogenation (nickel), gives menthone in almost quantitative yield; on oxidation with ferric chloride, thymol is obtained (Smith *et al.*, 1920). These reactions



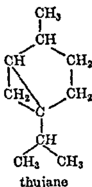
show that piperitone is *p*-menthene-3-one, but do not show the position of the double bond. This had been shown by Schimmel (1910), who found that on oxidation with alkaline permanganate, piperitone gave α-hydroxy-α-methyl-α′-*iso*propyladipic acid, II, γ-acetyl-α-*iso*propylbutyric acid, III, and α-*iso*propylglutaric acid,

IV. These results can be explained only if piperitone is *p*-menth-1-en-3-one, I. This structure for piperitone has been confirmed by various syntheses (*e.g.*, Henecka, 1948; Birch *et al.*, 1949).

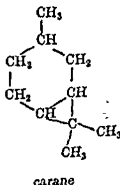
BICYCLIC MONOTERPENES

§19. Introduction. The bicyclic monoterpenes may be divided into three classes according to the size of the *second* ring, the first being a six-membered ring in each class.

Class I (6- + 3-membered ring).

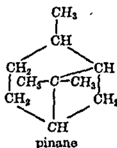


thujane



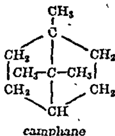
carane

Class II (6- + 4-membered ring).

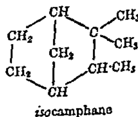


pinane

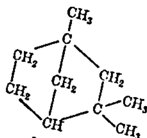
Class III (6- + 5-membered ring).



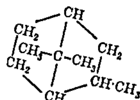
camphane



*iso*camphane



fenchane

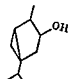


*iso*bornylane

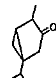
It is important to note that the two rings do not lie in one plane, but are almost perpendicular to each other.

§20. Thujone and its derivatives. The members of this group which occur naturally are the following:

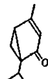

α-thujene



thujyl alcohol

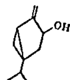


thujone



umbellulone


sabinene



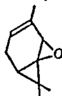
sabinol

§21. Carane and its derivatives. It appears that only three carane derivatives occur naturally:



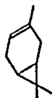
car-3-ene


car-4-ene

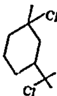


car-3-ene-5:6-epoxide

Car-3-ene occurs in Swedish pine needle oil. It is a liquid, b.p. 170°; when treated with hydrogen chloride it forms a mixture of sylvestrene dihydrochloride (see §15) and dipentene dihydrochloride (§13).











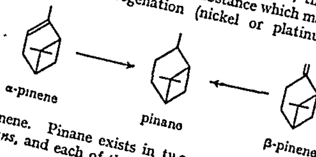
(+)-Car-4-ene, b.p. 165·5–167°/707 mm., occurs in various

first prepared by Baeyer *et al.* (1894) by the action of hydrogen bromide on dihydrocarvone, which was then treated with ethanolic potassium hydroxide, whereupon carone was obtained.

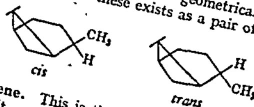
The structure of carone was established by Baeyer *et al.* (1896), who obtained caronic acid on oxidation of carone with permanganate. Baeyer suggested that caronic acid was a *cyclo*propane derivative, and this was confirmed by synthesis (Perkin, junior, and Thorpe, 1899), starting with ethyl β:β-dimethylacrylate and ethyl cyanoacetate (see opposite page).

An interesting point about carone is that its ultraviolet absorption spectrum shows similarities to that of α:β-unsaturated ketones (Klotz, 1941).

§22. Pinane and its derivatives. **Pinane**, the parent compound of this group, is a synthetic substance which may be prepared by the catalytic hydrogenation (nickel or platinum) of either



α- or β-pinene. Pinane exists in two geometrical isomeric forms, *cis* and *trans*, and each of these exists as a pair of enantiomorphs.



§22a. α-Pinene. This is the most important member of the pinane class. It occurs in both the (+)- and (−)-forms in all turpentine oils; it is a liquid, b.p. 156°.

The analytical evidence for the structure of α-pinene may conveniently be divided into two sections, each section leading independently to the structure, and the two taken together giving very powerful evidence for the structure assigned.

Method 1. The molecular formula of α-pinene is $C_{10}H_{16}$, and since α-pinene adds on two bromine atoms, one double bond is present in the molecule. Thus the parent hydrocarbon is $C_{10}H_{18}$, and since this corresponds to the general formula C_nH_{2n-2}, the general formula of compounds containing *two* rings, it therefore follows that α-pinene is bicyclic (Wallach, 1887–1891). In the

preparation of α-pinene nitrosochloride (by the action of nitrosyl chloride on α-pinene) the by-products which were formed were steam distilled, and the compound *pinol*, $C_{10}H_{16}O$, was thereby obtained. Pinol adds on one molecule of bromine to form pinol dibromide, and so pinol contains one double bond. Furthermore, the action of lead hydroxide on pinol dibromide converts the latter into pinol glycol, $C_{10}H_{16}O(OH)_2$, and this, on oxidation, gives terpenylic acid (Wallach *et al.*, 1889). Pinol (III) is also obtained by the action of sodium ethoxide on α-terpineol dibromide, II (Wallach, 1893). Wagner (1894) showed that the oxidation of pinol with permanganate gives pinol glycol (IV), which is further oxidised to terpenylic acid (V). All these facts can be explained as follows, based on I being the structure of α-terpineol (see also §11).

I —Br_2→ II —C_2H_5ONa→ III (or ...) —[O]→ IV —[O]→ V

Support for the structure given for pinol (III) is obtained from the fact that oxidation of *sobrerol* (pinol hydrate) produces a tetrahydric alcohol, *sobrerythritol*. Sobrerol itself is readily prepared by the action of hydrogen bromide on pinol, followed by sodium hydroxide. These reactions may thus be formulated:

pinol —HBr→ pinol hydrobromide —NaOH→ sobrerol —$KMnO_4$→ sobrerythritol

Thus, if the formula for α-pinene is VI, then the formation of the above substances can be explained. This structure also accounts for other reactions of α-pinene, *e.g.*, its ready hydration to α-terpineol (see later).



VI

Although the Wagner formula (VI) for α-pinene readily explains all the facts, there is no *direct* evidence for the existence of the *cyclo*butane ring. Such evidence was supplied by Baeyer (1896). This is described in method 2.

Method 2. As in *method* 1, α-pinene was shown to be bicyclic. When treated with ethanolic sulphuric acid, α-pinene is converted into α-terpineol (Flavitzky, 1879). Therefore α-pinene contains a six-membered ring and another ring (since it is bicyclic), the carbon skeleton of pinene being such as to give α-terpineol when this second ring opens. Since, in the formation of α-terpineol, one molecule of water is taken up and the hydroxyl group becomes attached to C_8, this suggests that the C_8 of α-terpineol is involved in forming the second ring in α-pinene. There are three possible points of union for this C_8, resulting in two three-membered and one four-membered ring (see VII); at the same time the position of the double bond in α-pinene is also shown by the conversion into α-terpineol (I).



VII





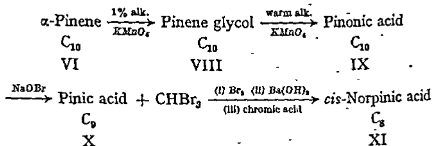
I



VII*a*

A point of interest here is that there are actually *four* possible points of union for C_8, the three shown in VII and the fourth being at the double bond to form a four-membered ring (VII*a*). This one, however, was rejected on the grounds of *Bredt's rule* (1924) which states that a double bond cannot be formed by a carbon atom occupying the bridge-head (of a bicyclic system). The explanation for this rule is that structures such as VII*a* have a large amount of strain.

This second ring was shown to be four-membered by Baeyer (1896), who carried out the following series of reactions.

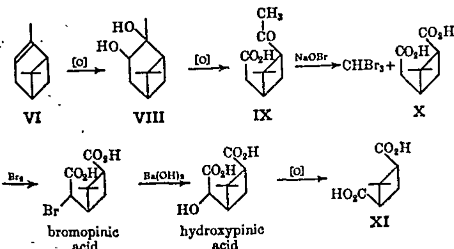


Pinene glycol, $C_{10}H_{16}(OH)_2$, is produced by hydroxylation of the double bond in α-pinene, and pinonic acid, $C_{10}H_{16}O_3$, is produced by scission of the glycol bond. Pinonic acid was shown to be a saturated keto-monocarboxylic acid. The formation of pinic acid, $C_9H_{14}O_4$, and bromoform, indicates the presence of an acetyl group in pinonic acid. Pinic acid, which was shown to be a saturated dicarboxylic acid, on treatment with bromine, then barium hydroxide, and finally the product oxidised with chromic acid, gives *cis*-norpinic acid, $C_8H_{12}O_4$. This was shown to be a saturated dicarboxylic acid, and so its formula may be written $C_6H_{10}(CO_2H)_2$. Furthermore, since α-pinene contains two methyl groups attached to a carbon atom in the second ring (see VII), and it is the *other* ring (the six-membered one containing the double bond) that has been opened by the above oxidation, then norpinic acid (with this second ring intact) contains these two methyl groups. Thus the formula for norpinic acid may be written $(CH_3)_2C_4H_4(CO_2H)_2$. Hence, regarding the methyl and carboxyl groups as substituents, ... acid is [illegible] basis, pinic acid could therefore be a *cyclo*butane derivative with one side-chain of $—CH_2·CO_2H$.

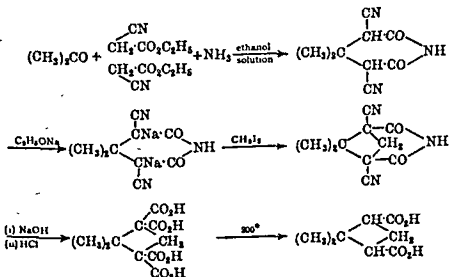
Baeyer therefore *assumed* that pinic and norpinic acids contained a *cyclo*butane ring, and so suggested the following structures to account for the above reactions, accepting structure VI for α-pinene, the structure already proposed by Wagner (1894).

The synthesis of norpinic acid (*to confirm the above reactions*) proved to be a very difficult problem, and it was not carried out until 1929, when Kerr succeeded with the following ingenious method (apparently the presence of the *gem* dimethyl group prevents closure to form the *cyclo*butane ring).

The norpinic acid obtained was the *trans*-isomer; this is readily

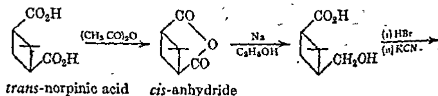


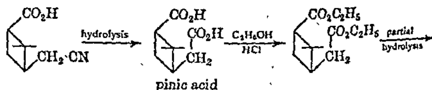
converted into the *cis*-isomer (the isomer obtained from the oxidation of α-pinene) by heating the *trans* acid with acetic anhydride, whereupon the *cis* anhydride is formed and this, on hydrolysis, gives the *cis* acid (Simonsen *et al.*, 1929).

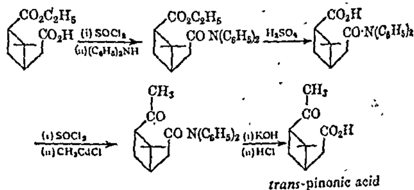


The total synthesis of α-pinene has now been carried out in the following way. Guha *et al.* (1937) synthesised pinic acid from norpinic acid, and Rao (1943) synthesised pinonic acid from synthetic pinic acid (see next page).

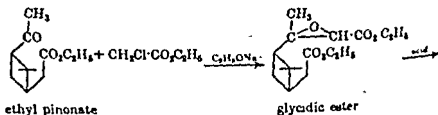
Ruzicka *et al.* (1920–1924) had already synthesised α-pinene starting from pinonic acid (obtained by the oxidation of α-pinene). Thus we now have a total synthesis of α-pinene. Ruzicka's synthesis

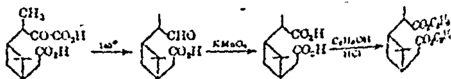


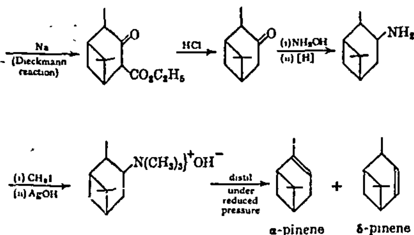




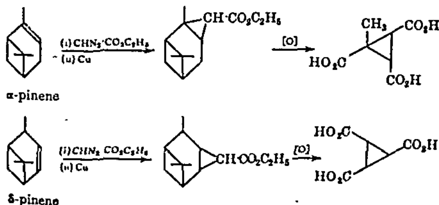
makes use of the Darzens glycidic ester synthesis (see Vol. I) the steps are:







The final step gives a mixture of two compounds, α- and δ-pinene The former was identified by the preparation of the nitrosochloride; this proves that one of the products is α-pinene, but does not prove which is α and which is δ. These are differentiated by consideration of the analytical evidence; the following evidence also supports the structure given for α-pinene. This evidence is based on the fact that diazoacetic ester combines with compounds containing a double bond to form pyrazoline derivatives, and these, on heating alone or with copper powder, decompose to produce *cyclo*propane derivatives (see also §2a. XII). When the two pinenes were subjected to this treatment, and the resulting compounds oxidised,

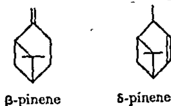


α-pinene gave 1-methyl*cyclo*propane-1:2.3-tricarboxylic acid, and δ-pinene *cyclo*propane-1:2.3-tricarboxylic acid. These products are in accord with the structures assigned to α- and δ-pinene.

Examination of the α-pinene structure shows that two dissimilar

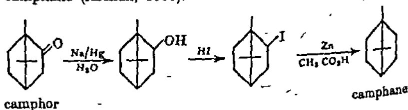
asymmetric carbon atoms are present; thus two pairs of enantiomorphs are possible. In practice, however, only one pair is known. This is due to the fact that the four-membered ring can only be fused to the six-membered one in the *cis*-position; *trans* fusion is impossible. Thus only the enantiomorphs of the *cis*-isomer are known.

Isomeric with α-pinene are β- and δ-pinene; the former occurs naturally, the latter is synthetic (see Ruzicka's synthesis)



§23. Camphane and its derivatives. Camphane, $C_{10}H_{18}$, is a synthetic compound, and may be prepared from camphor, *e.g.*,

(i) By reduction of camphor to a mixture of borneols (§23b), these then converted to the bornyl iodides which are finally reduced to camphane (Aschan, 1900).



(ii) Camphor may also be converted into camphane by means of the Wolff–Kishner reduction (see also Vol. I).



Camphane is a solid, m.p. 156°; it is optically inactive.

§23a. Camphor. This occurs in nature in the camphor tree of Formosa and Japan. It is a solid, m.p. 179°, and is optically active; the (+)- and (−)-forms occur naturally, and so does racemic camphor, which is the usual form of synthetic camphor (from α-pinene; see later).

A tremendous amount of work was done before the structure of camphor was successfully elucidated; in the following account only a small part of the work is described, but it is sufficient to justify the structure assigned to camphor.

The molecular formula of camphor is $C_{10}H_{16}O$, and the general reactions and molecular refractivity of camphor show that it is saturated. The functional nature of the oxygen atom was shown to be oxo by the fact that camphor formed an oxime, etc., and that it was a keto group was deduced from the fact that oxidation of camphor gives a *dicarboxylic acid containing 10 carbon atoms*; a *monocarboxylic* acid containing 10 carbon atoms cannot be obtained (this type of acid would be expected if camphor contained an *aldehyde* group). From the foregoing facts it can be seen that the parent hydrocarbon of camphor has the molecular formula $C_{10}H_{18}$; this corresponds to C_nH_{2n-2}, and so camphor is therefore bicyclic. Camphor contains a $—CH_2·CO—$ group, since it forms an oxime with nitrous acid (*iso*amyl nitrite and hydrogen chloride). Finally, distillation of camphor with zinc chloride or phosphorus pentoxide produces *p*-cymene.

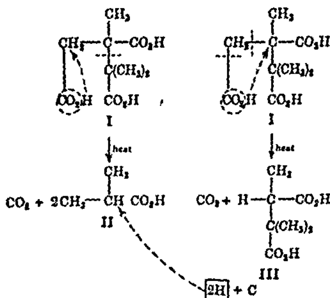
Bredt (1893) was the first to assign the correct formula to camphor (over 30 have been proposed). Bredt based his formula on the above facts and also on the facts that (*a*) oxidation of camphor with nitric acid gives **camphoric acid**, $C_{10}H_{16}O_4$ (Malaguti, 1837); (*b*) oxidation of camphoric acid (or camphor) with nitric acid gives **camphoronic acid**, $C_9H_{14}O_6$ (Bredt, 1893).

[illegible]

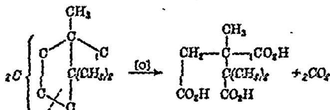
tivity showed that it is saturated. Thus, in the formation of camphoric acid from camphor, the ring containing the keto group is opened, and consequently camphoric acid must be a monocyclic compound.

Camphoronic acid was shown to be a saturated tricarboxylic acid, and on distillation at atmospheric pressure, it gave *iso*butyric acid, II, *trimethylsuccinic acid*, III, *carbon dioxide and carbon* (and a small amount of some other products). Bredt (1893) therefore *suggested* that camphoronic acid is $\alpha{:}\alpha{:}\beta$-trimethyltricarballylic acid, I, since this structure would give the required decomposition products. In the following equations, the left-hand side molecule is imagined to break up as shown; one molecule of carbon dioxide and two molecules of *iso*butyric acid are produced (but there is a shortage of two hydrogen atoms). The right-hand side molecule breaks up to form one molecule of trimethylsuccinic acid, one molecule of carbon dioxide, one atom of carbon, and *two atoms of*

hydrogen which now make up the shortage of the left-hand side molecule. Thus:

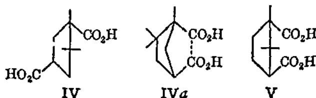


Hence, if camphoronic acid has structure I, then camphoric acid (and camphor) must contain *three methyl groups*. On this basis, the formula of camphoric acid, $C_{10}H_{16}O_4$, can be written as $(CH_3)_3C_5H_5(CO_2H)_2$. The parent (saturated) hydrocarbon of this is C_5H_{10}, which corresponds to C_nH_{2n}, *i.e.*, camphoric acid is a *cyclo*pentane derivative (this agrees with the previous evidence that camphoric acid is monocyclic). Thus the oxidation of camphoric acid to camphoronic acid may be written:

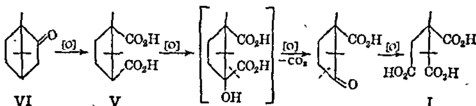


This skeleton, plus *one* carbon atom, arranged with two carboxyl groups, will therefore be the structure of camphoric acid. Now camphoric anhydride forms only one monobromo derivative (bromine and phosphorus); therefore there is only *one* α-hydrogen atom in camphoric acid. Thus the carbon atom of one carboxyl group must be $_1C$ (this is the only carbon atom joined to a tertiary carbon atom). Furthermore, $_1C$ must be the carbon of the keto or methylene group in camphor, since it is these two groups which produce the two carboxyl groups in camphoric acid. The problem

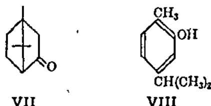
is now to find the position of the other carboxyl group in camphoric acid. Its position must be such that when the *cyclo*pentane ring is opened to give camphoronic acid, one carbon atom is readily lost. Using this as a working hypothesis, then there are only two reasonable structures for camphoric acid, IV and V. IV may be rewritten



as IV*a*, and since the two carboxyl groups are produced from the —CH_2·CO— group in camphor, the precursor of IV*a* (*i.e.*, camphor) will contain a six-membered ring with a *gem*-dimethyl group. This structure cannot account for the conversion of camphor into *p*-cymene. On the other hand, V accounts for all the facts given in the foregoing discussion. Bredt therefore assumed that V was the structure of camphoric acid, and that VI was the structure of camphor, and proposed the following reactions to show the relationships between camphor, camphoric acid and camphoronic acid.



Bredt, however, realised that if camphor had structure VII, then all the foregoing facts would be equally satisfied, but he rejected VII in favour of VI for a number of reasons. One simple fact that

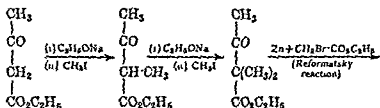


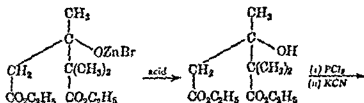
may be used here for rejection of VII is that camphor gives carvacrol, VIII, when distilled with iodine. The formation of this compound can be expected from VI but not from VII.

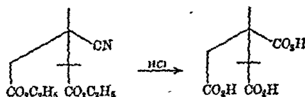
Formula VI for camphor was accepted with reserve at the time when Bredt proposed it (in 1893), but by 1903 all the deductions

of Bredt were confirmed by the syntheses of camphoronic acid, camphoric acid and camphor.

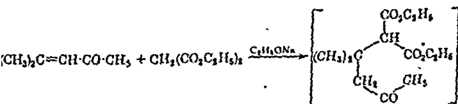
Synthesis of (±)-camphoronic acid (Perkin, junior, and Thorpe, 1897).

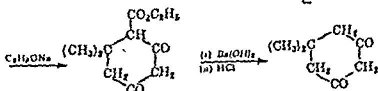


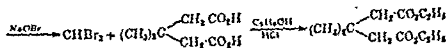




Synthesis of (±)-camphoric acid (Komppa, 1903). Komppa (1899) first synthesised β·β-dimethylglutaric ester as follows, start-

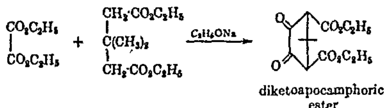


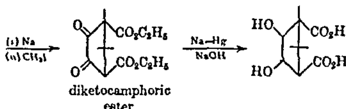


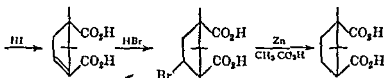


ing with mesityl oxide and ethyl malonate. The product obtained was 6.6-dimethyl*cyclo*hexane-2:4-dione-1-carboxylic ester (this is produced first by a Michael condensation, followed by a Dieckmann reaction). On hydrolysis, followed by oxidation with sodium hypobromite, β:β-dimethylglutaric acid was obtained (*cf.* carone, §21).

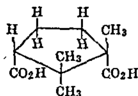
Komppa (1903) then prepared camphoric acid as follows:



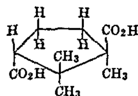




The structure given for camphoric acid can exist in two geometrical isomeric forms, *cis* and *trans*, neither of which has any elements of symmetry. Thus four optically active forms are possible; all are known, and correspond to the (+)- and (−)-forms of camphoric acid and *iso*camphoric acid. Since camphoric acid forms an anhydride, and *iso*camphoric acid does not, the former is the *cis*-isomer, and the latter the *trans*- (§5i. IV).



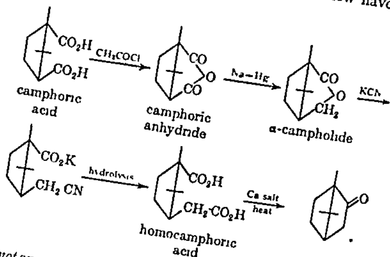
camphoric acid, m.p. 187°



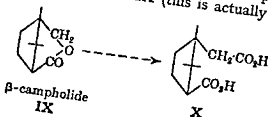
iso-camphoric acid, m.p. 171-172°

Synthesis of camphor (Haller, 1896). Haller started with camphoric acid prepared by the oxidation of camphor, but since

the acid was synthesised later by Komppa, we now have a total synthesis of camphor.



This is *not* an unambiguous synthesis, since the campholide obtained might have had the structure IX (this is actually β-campholide).



In this case, homocamphoric acid would have had structure X, and this would have given camphor with structure VII which, as we have seen, was rejected.

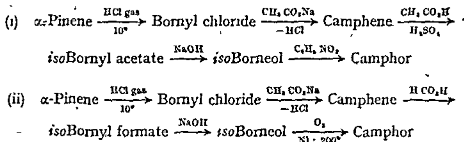
Stereochemistry of camphor. Camphor has two dissimilar asymmetric carbon atoms (the same two as in camphoric acid), but only one pair of enantiomorphs is known. This is due to the fact that only the *cis*-form is possible; *trans* fusion of the *gem*-dimethylmethylene bridge to the *cyclo*hexane ring is impossible. Thus only the enantiomorphs of the *cis*-isomer are known (*cf.* α-pinene, §22a).

Some derivatives of camphor. The positions of substituent groups in camphor are indicated by numbers or by the Greek letters α (= 3), β or ω (= 10), and π (= 8 or 9). When (+)-camphor is heated with bromine at 100°, α-bromo-(+)-camphor is produced. This, on warming with sulphuric acid, is converted into α-bromo-(+)-camphor-π-sulphonic acid which, on reduction, forms (+)-camphor-π-sulphonic acid. (±)-Camphor-π-sulphonic acid is ob-

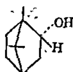


tained by the sulphonation of (+)-camphor with fuming sulphuric acid; under these conditions, (+)-camphor is racemised. On the other hand, sulphonation of (+)-camphor with sulphuric acid in acetic anhydride solution produces (+)-camphor-β-sulphonic acid. These various (+)-camphorsulphonic acids are very valuable reagents for resolving racemic bases (§10 iv. II).

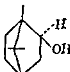
Commercial preparation of camphor. Synthetic camphor is usually obtained as the racemic modification. The starting material is α-pinene, and the formation of camphor involves the Wagner-Meerwein rearrangements (see §23d). Scheme (i) is the earlier method, and (ii) is the one that is mainly used now



§23b. **Borneols,** $C_{10}H_{18}O$. There are two stereoisomeric compounds of the formula $C_{10}H_{18}O$; these correspond to **borneol** and ***iso*borneol,** and both are known in the (+)- and (−)-forms. The borneols occur widely distributed in essential oils, but it appears that the *iso*borneols have been isolated from only one essential oil. Borneol and *iso*borneol are secondary alcohols, and the evidence now appears to be conclusive that borneol has the *endo*-configuration in which the *gem*-dimethyl bridge is above the plane of the *cyclo*hexane ring and the hydroxyl group is below the plane. *iso*Borneol



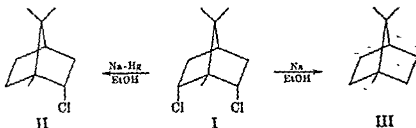
borneol
m.p. 208·5°



*iso*borneol
m.p. 217°

has the *exo*-configuration in which the bridge and the hydroxyl group are both above the plane of the *cyclo*hexane ring.

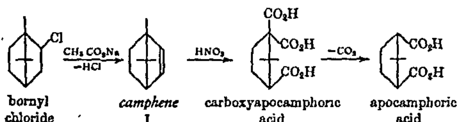
Kwart *et al.* (1956) have now obtained direct evidence on the configuration of bornyl chloride. Bornyl dichloride (I), the structure of which has been established by Kwart (1953), is converted into bornyl chloride (II) by sodium amalgam and ethanol, and into camphane (III) by sodium and ethanol.



Both borneol and *iso*borneol are produced when camphor is reduced, but the relative amounts of each are influenced by the nature of the reducing agent used, *e.g.*, electrolytic reduction gives mainly borneol, whereas catalytic hydrogenation (platinum) gives mainly *iso*borneol; *iso*borneol is also the main product when aluminium *iso*propoxide is used as the reducing agent (the Meerwein–Ponndorf–Verley reduction; see Vol. I). Borneol is converted into a mixture of bornyl and *iso*bornyl chlorides by the action of phosphorus pentachloride. Borneol and *iso*borneol are both dehydrated to camphene (§23c), but the dehydration occurs more readily with *iso*borneol than with borneol. Both alcohols are oxidised to camphor, but whereas borneol can be dehydrogenated to camphor by means of a copper catalyst, *iso*borneol cannot.

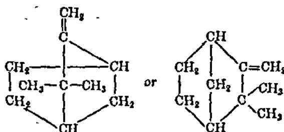
§23c. Camphene and Bornylene. Camphene, $C_{10}H_{16}$, m.p. 51–52°, occurs naturally in the (+)-, (−)-, and (±)-forms. It may be prepared by the removal of a molecule of hydrogen chloride from bornyl and *iso*bornyl chlorides by means of sodium acetate, or by the dehydration of the borneols with potassium hydrogen sulphate. These methods of preparation suggest that camphene contains a double bond, and this is supported by the fact that camphene adds on one molecule of bromine or one molecule of hydrogen chloride. Oxidation of camphene with dilute nitric acid produces carboxyapocamphoric acid, $C_{10}H_{14}O_6$, and apocamphoric acid, $C_9H_{14}O_4$ (Marsh *et al.*, 1891). The formation of the former

acid, which contains the same number of carbon atoms as camphene, implies that the double bond in camphene is in a ring; and the fact that carboxyapocamphoric acid is converted into apocamphoric acid when heated above its melting point implies that the former contains two carboxyl groups attached to the same carbon atom (*cf.* malonic ester syntheses). These facts were explained by giving



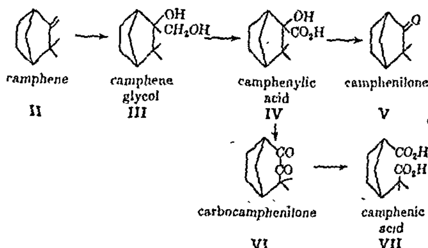
camphene the formula shown (I). The structure of apocamphoric acid was later proved by synthesis (Komppa, 1901; *cf.* camphoric acid, §23a).

This structure for camphene, however, was opposed by Wagner. The oxidation of camphene with dilute permanganate gives camphene glycol, $C_{10}H_{16}(OH)_2$ [Wagner, 1890]. This glycol is saturated, and so camphene is a bicyclic compound (so, of course, is structure I). On further oxidation of camphene glycol, Wagner (1896, 1897) obtained camphenic acid, $C_{10}H_{16}O_4$ (a dibasic acid), and camphenylic acid, $C_{10}H_{16}O_3$ (a hydroxy-monobasic acid), which, on oxidation with lead dioxide, gave camphenilone, $C_9H_{14}O$ (a ketone). According to Wagner, it was difficult to explain the formation of these compounds if camphene had structure I. Wagner (1899) therefore suggested that camphene is formed by a molecular rearrangement when the borneols or bornyl chlorides are converted into camphene, and proposed structure II for camphene (see also §23d).



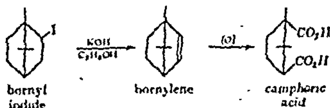
II

With this formula, the formation of camphene glycol, camphenylic acid and camphenilone could be explained as follows:



Although it was easy to explain the formation of III, IV and V, it was difficult to explain the formation of VII. The formation of VII was explained by later workers, who suggested it was produced *via* carbocamphenilone, VI. Another difficulty of the camphene formula, II, is that it does not explain the formation of apocamphoric acid when camphene is oxidised with nitric acid (see above). The course of its formation has been suggested by Komppa (1908, 1911), who proposed a mechanism involving a Wagner rearrangement.

Structure II for camphene is supported by the fact that treatment of bornyl iodide with ethanolic potassium hydroxide at 170° gives **bornylene**, $C_{10}H_{16}$ (m.p. 98°), as well as camphene (Wagner *et al.*, 1899). Bornylene is readily oxidised by permanganate to camphoric acid; it therefore follows that bornylene has the structure I, the structure originally assigned to camphene; no rearrangement occurs in the formation of bornylene.



Ozonolysis of camphene gives camphenilone and formaldehyde (Harries *et al.*, 1910); these products are in keeping with the Wagner formula for camphene.

$$\text{II} \xrightarrow{O_3} \text{V} + CH_2O$$

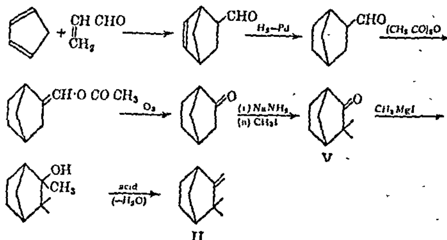
Further support for this structure for camphene is afforded by the work of Buchner *et al.* (1913). These workers showed that camphene reacts with diazoacetic ester, and when the product is hydrolysed and then oxidised, *cyclo*propane-1:1:2-tricarboxylic

$$\text{II} + CHN_2{\cdot}CO_2C_2H_5 \longrightarrow \text{(adduct)}{-}CO_2C_2H_5 \xrightarrow[\text{(ii) oxidation}]{\text{(i) hydrolysis}} \text{VIII}$$

acid, VIII, is produced. VIII is to be expected from structure II, but not from I; I (bornylene) would give *cyclo*propane-1:2:3-tricarboxylic acid, IX.

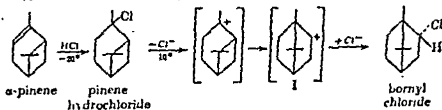
$$\text{I} + CHN_2{\cdot}CO_2C_2H_5 \longrightarrow \text{(adduct)}{-}CO_2C_2H_5 \xrightarrow[\text{(ii) oxidation}]{\text{(i) hydrolysis}} \text{IX}$$

Lipp (1914) has synthesised camphenic acid (VII), and showed that it has the structure assigned to it by Wagner. Finally, camphene has been synthesised as follows (Diels and Alder, 1928–1931).

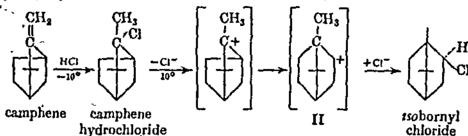


§23d. **Wagner-Meerwein rearrangements.** Wagner, as we have seen, proposed a molecular rearrangement to explain the formation of camphene from the borneols and bornyl chlorides. Wagner also recognised that a molecular rearrangement occurred when α-pinene was converted into bornyl chloride. Many other investigations concerning rearrangements in the terpene field were carried out by Meerwein and his co-workers, *e.g.*, when α-pinene is treated in ethereal solution at $-20°$ with hydrogen chloride, the product is pinene hydrochloride. This is unstable, and if the temperature is allowed to rise to about 10°, the pinene hydrochloride rearranges to bornyl chloride (Meerwein *et al.*, 1922). Rearrangements such as these which occur with bicyclic monoterpenes are known as *Wagner–Meerwein rearrangements.* They are all believed to take place *via* the formation of a carbonium ion, but whether this ion is actually free is still uncertain. The problem is also complicated by the steric course of the rearrangement, *i.e.*, whether a Walden inversion occurs or not. The following gives a simplified version of the mechanism of a number of cases involving the Wagner–Meerwein rearrangement; and it should be noted that the mechanism given is a special case of the Whitmore mechanism (see §2h. VI).

(i) *The conversion of α-pinene into bornyl chloride.*

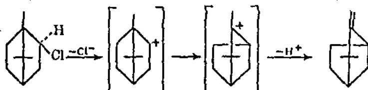


(ii) *The conversion of camphene into* isobornyl *chloride.*

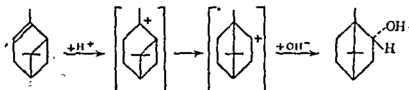


It will be seen that carbonium ion I in example (i) appears to be identical with carbonium ion II in this example, yet the former gives bornyl chloride, and the latter *iso*bornyl chloride. The reason for this is uncertain.

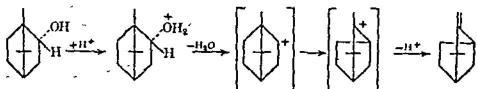
(iii) *Dehydrohalogenation of* isobornyl *chloride to camphene.*



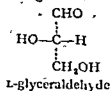
(iv) *Hydration of α-pinene to borneol (with acids).*



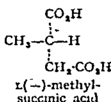
(v) *Dehydration of borneol to camphene (with acids).*



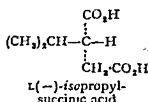
§23e. **Correlation of configurations of terpenes.** This has been made possible by the work of Fredga on quasi-racemic compounds (see §9a. II). This author has established the following configurations:



L-glyceraldehyde

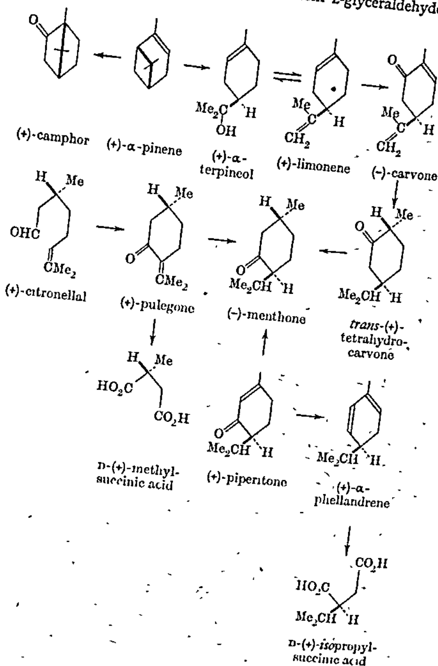


L(−)-methylsuccinic acid



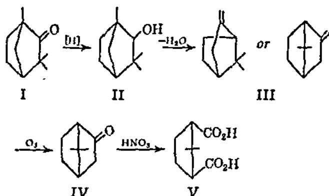
L(−)-*iso*propylsuccinic acid

By means of these configurations, combined with various interrelations obtained by oxidative degradations and by molecular rearrangements, it has been possible to correlate the configurations of many mono- and bicyclic terpenes with L-glyceraldehyde, *e.g.*,

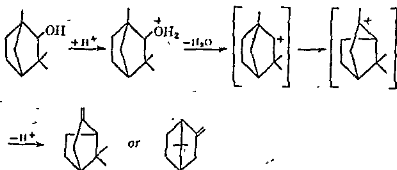


§24. Fenchane and its derivatives. The most important natural terpene of this group is **fenchone**; this occurs in oil of fennel. It is a liquid, b.p. 192–193°, and is optically active, both enantiomorphs occurring naturally.

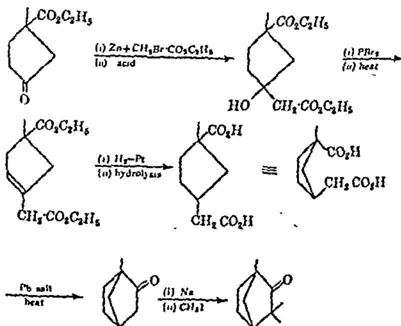
The molecular formula of fenchone is $C_{10}H_{16}O$, and the compound behaves as a ketone. When fenchone (I) is reduced with sodium and ethanol, fenchyl alcohol, $C_{10}H_{18}O$ (II), is produced, and this, on dehydration under the influence of acids, gives α-fenchene, $C_{10}H_{16}$ (III). On ozonolysis, α-fenchene is converted into α-fenchocamphorone, $C_9H_{14}O$ (IV), which, on oxidation with nitric acid, forms apocamphoric acid, V, a compound of known structure. This work was carried out by Wallach *et al.* (1890–1898), but it was Semmler (1905) who was the first to assign the correct structure to fenchone; the foregoing reactions may be formulated:



It should be noted that the dehydration of fenchyl alcohol, II, to α-fenchene, III, occurs *via* a Wagner–Meerwein rearrangement; the mechanism for this reaction may thus be written (cf. §23d):



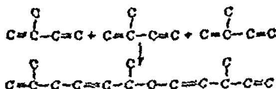
The structure of fenchone has been confirmed by synthesis (Ruzicka, 1917).



SESQUITERPENES

§25. **Introduction.** The sesquiterpenes, in general, form the higher boiling fraction of the essential oils; this provides their chief source. Wallach (1887) was the first to suggest that the sesquiterpene structure is built up of three isoprene units; this has been shown to be the case for the majority of the known sesquiterpenes, but there are some exceptions.

The sesquiterpenes are classified into four groups according to the number of rings present in the structure. If we use the *isoprene rule*, then when three isoprene units are linked (head to tail) to form an acyclic sesquiterpene hydrocarbon, the latter will contain *four* double bonds. Each isoprene unit contains *two* double bonds, but one disappears for each pair that is connected:



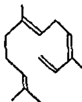
When this open-chain compound is converted into a monocyclic structure, another double bond is utilised in the process, and so monocyclic sesquiterpene hydrocarbons contain three double bonds. In a similar manner, it will be found that a bicyclic structure contains two double bonds, and a tricyclic one. Thus the nature of the sesquiterpene skeleton is also characterised by the number of double bonds present in the molecule. The sesquiterpene hydrocarbon structures may also be distinguished by the calculation of the molecular refractivities for the various types of structures, and then using these values to help elucidate the structures of new sesquiterpenes; *e.g.*, zingiberene (§27a).

Class of sesquiterpene	Number of double bonds	Molecular refractivity
Acyclic	4	69 5
Monocyclic . . .	3	67·8
Bicyclic	2	66·1
Tricyclic	1	64·4

This type of information can also be used with the monoterpenes, but in this case it has not been so useful as in the sesquiterpenes.

ACYCLIC SESQUITERPENES

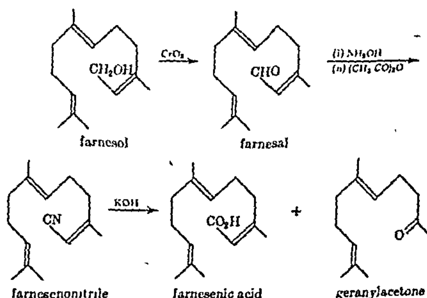
§26. **Farnesene,** $C_{15}H_{24}$, b.p. 128–130°/12 mm., is obtained by the dehydration of farnesol with potassium hydrogen sulphate (Harries *et al.*, 1913).



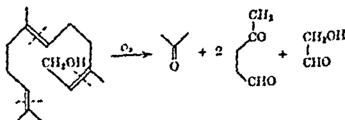
α-farnesene

§26a. **Farnesol,** $C_{15}H_{26}O$, b.p. 120°/0·3 mm., occurs in the oil of ambrette seeds, etc. Its structure was elucidated by Kerschbaum (1913) as follows. When oxidised with chromic acid, farnesol is converted into farnesal, $C_{15}H_{24}O$, a compound which behaves as an aldehyde. Thus farnesol is a primary alcohol. Conversion of farnesal into its oxime, followed by dehydration with acetic

anhydride, produces a cyanide which, on hydrolysis with alkali, forms farnesenic acid, $C_{15}H_{24}O_2$, and a ketone, $C_{13}H_{22}O$. This ketone was then found to be dihydro-*pseudo*-ionone (geranylacetone). In the formation of this ketone, two carbon atoms are removed from its precursor. This reaction is characteristic of $\alpha\beta$-unsaturated carbonyl compounds, and so it is inferred that the precursor, farnesenic acid (or its nitrile), is an $\alpha\beta$-unsaturated compound. Thus the foregoing facts may be formulated as follows, on the basis of the known structure of geranylacetone.

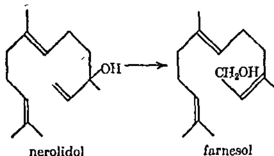


Kerschbaum's formula has been confirmed by Harries *et al.* (1913), who obtained acetone, lævulaldehyde and glycolaldehyde on the ozonolysis of farnesol.

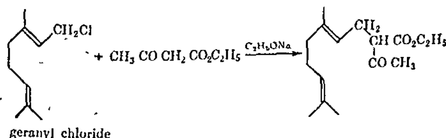


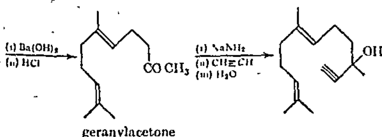
Ozonolysis, however, also gave some formaldehyde, thus indicating the presence of the isopropenyl end-group as well as the *isopropylidene* end-group (but *cf.* citral, §5). Ruzicka (1923) synthesised

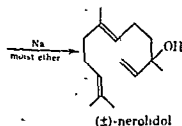
farnesol (with the *iso*propylidene end-group) by the action of acetic anhydride on synthetic nerolidol (*cf.* linalool, §8).



§26b. Nerolidol, $C_{15}H_{26}O$, b.p. 125–127°/4·5 mm., occurs in the oil of neroli, etc., in the (+)-form. Nerolidol is isomeric with farnesol, and Ruzicka (1923) showed that the relationship between the two is the same as that between linalool and geraniol (see §8). Ruzicka (1923) confirmed the structure of nerolidol by synthesis.



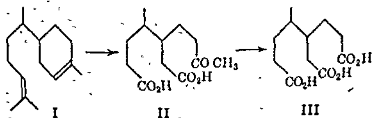




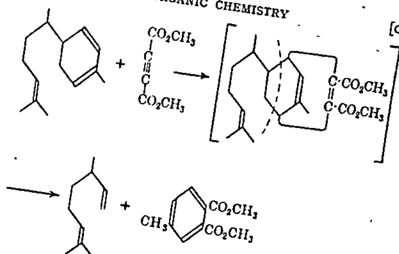
of succinic acid. These products are readily accounted for by III; and this structure has been confirmed by synthesis (Ruzicka *et al.*, 1932).

§27a. Zingiberene, $C_{15}H_{24}$, b.p. 134°/14 mm., occurs in the (−)-form in ginger oil. It forms a dihydrochloride with hydrogen chloride, and thus apparently contains two double bonds. The molecular-refractivity, however, indicates the presence of three double bonds and, if this be the case, zingiberene is monocyclic (see §25). The presence of these three double bonds is conclusively shown by the fact that catalytic hydrogenation (platinum) converts zingiberene into hexahydrozingiberene, $C_{15}H_{30}$. Zingiberene can be reduced by means of sodium and ethanol to dihydrozingiberene, $C_{15}H_{26}$; this indicates that two of the double bonds are probably conjugated (Semmler *et al.*, 1913). Further evidence for this conjugation is afforded by the fact that zingiberene shows optical exaltation, whereas dihydrozingiberene does not. The absorption spectrum of zingiberene also shows the presence of conjugated double bonds (Gillam *et al.*, 1940).

Ozonolysis of zingiberene gives acetone, lævulic acid and succinic acid (Ruzicka *et al.*, 1929). Since these products are also obtained from bisabolene (§27), it appears probable that zingiberene and bisabolene have the same carbon skeleton. Oxidation of dihydrozingiberene, I, with permanganate gives a keto-dicarb- [illegible] ation with sodium hypo- [illegible] O_6 (III). Thus II must [illegible], and so, if I be assumed as the structure of dihydrozingiberene, the foregoing oxidation reactions may be formulated:

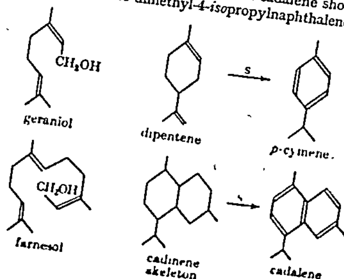


Thus I, with another double bond in conjugation with one already present, will be (probably) the structure of zingiberene. The position of this third double bond was shown as follows (Eschenmoser *et al.*, 1950). Zingiberene forms an adduct with methyl acetylenedicarboxylate, and this adduct (which was not isolated), on pyrolysis, gives 2,6-dimethylocta-2,7-diene and methyl 4-methylphthalate. These reactions can be explained on the assumption that zingiberene has the structure shown below.

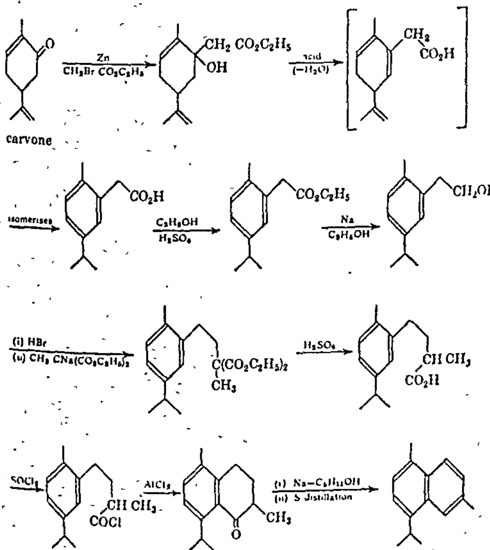


BICYCLIC SESQUITERPENES

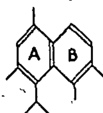
§28. Cadinene, $C_{15}H_{24}$, b.p. 134–136°/11 mm., occurs in the (−)-form in oil of cubebs, etc. Catalytic hydrogenation converts cadinene into tetrahydrocadinene, $C_{15}H_{28}$. Thus cadinene contains two double bonds and is bicyclic. On dehydrogenation with sulphur, cadinene forms cadalene, $C_{15}H_{18}$ (Ruzicka *et al.*, 1921). Cadalene does not add on bromine, and forms a picrate. This led to the belief that cadalene was an aromatic compound, and its structure was deduced as follows. Ruzicka assumed that the relationship of farnesol (§26a) to cadinene was analogous to that of geraniol (§7) to dipentene (§13). Furthermore, since dipentene gives *p*-cymene when dehydrogenated with sulphur, then cadalene should be, if the analogy is correct, 1:6-dimethyl-4-*iso*propylnaphthalene; thus:



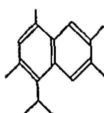
1:6-Dimethyl-4-*iso*propylnaphthalene was synthesised by Ruzicka *et al.* (1922), and was found to be identical with cadalene.



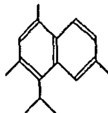
Thus cadinene has the carbon skeleton assumed. The only remaining problem is to ascertain the positions of the two double bonds in cadinene. Since the molecular refractivity shows no optical exaltation, the two double bonds are not conjugated (§11, I); this is supported by the fact that cadinene is not reduced by sodium and amyl alcohol. Ozonolysis of cadinene produces a compound containing the *same* number of carbon atoms as cadinene. The two double bonds are therefore in ring systems; but they cannot be in the *same* ring, since in this case carbon would have been lost on



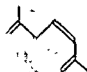
III

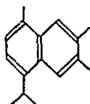


IV

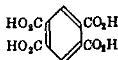


V



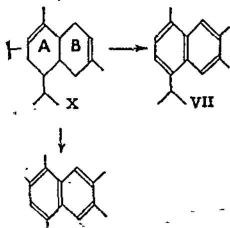


VII

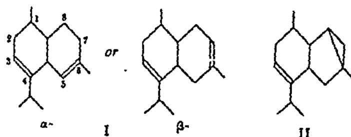


VIII

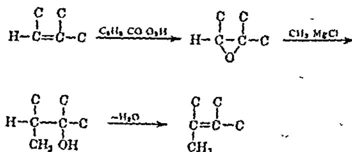
... of dimethylcadalene, but IV, with the two ... 6 and 7 in ring B, could give VIII. ... bond in cadinene in ring B is 6:7. From this ... eliminated. If the double bond in ring A ... then dimethylcadalene is IV, and mono- ... VII. Campbell and Soffer synthesised IV ... each was different from the methyl- ... from cadinene. Thus IV and VII ... the double bond in ring A cannot be ... which could give VIII ... was synthesised, and was found to be ... from cadinene. Cadinene ... the introduction of one or two methyl ... as follows:



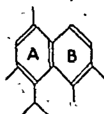
ozonolysis. Ruzicka *et al.* (1924) were thus led to suggest I (α or β) for the structure of cadinene, basing it on the relationship of cadinene to copaene, which had been given structure II by Semmler (1914). I was proposed mainly on the fact that copaene adds two



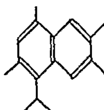
molecules of hydrogen chloride to form copaene dihydrochloride, [illegible] ne α and [illegible]. Struc- [illegible] Campbell and Soffer re-investigated the problem. These authors converted cadinene into its monoxide and dioxide by means of perbenzoic acid, treated these oxides with excess of methylmagnesium chloride, and then dehydrogenated the product with selenium. By this means, Campbell and Soffer obtained a monomethylcadalene from cadinene monoxide, and a dimethylcadalene from cadinene dioxide. Now the introduction of a methyl group *via* the oxide takes place according to the following scheme:



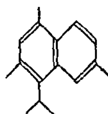
Thus the positions of the additional methyl groups show the positions of the double bonds in cadinene. The Ruzicka formula for cadinene would give dimethylcadalene III (from the α isomer) or IV (from the β), and the monomethylcadalenes would be V (from α or β), VI (from α) and VII (from β). Campbell and Soffer oxidised their dimethylcadalene, first with chromic acid and then with nitric acid, and thereby obtained pyromellitic acid (benzene-1:2.4:5-tetracarboxylic acid), VIII. The formation of VIII therefore rules



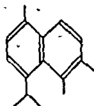
III



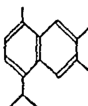
IV



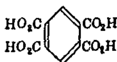
V



VI

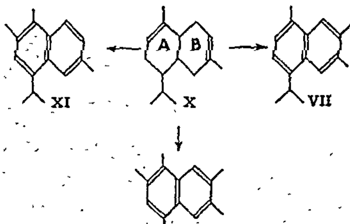


VII



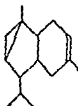
VIII

out III as the structure of dimethylcadalene, but IV, with the two methyl groups at positions 6 and 7 in ring B, could give VIII. Therefore the double bond in cadinene in ring B is 6:7. From this it follows that VI is also eliminated. If the double bond in ring A is as in structure I, then dimethylcadalene is IV, and monomethylcadalene is V or VII. Campbell and Soffer synthesised IV and VII, and found that each was different from the methylcadalenes they had obtained from cadinene. Thus IV and VII are incorrect; consequently the double bond in ring A cannot be 3:4. The only other dimethylcadalene which could give VIII on oxidation is IX. This was synthesised, and was found to be identical with the dimethylcadalene from cadinene. Cadinene must therefore be X, and the introduction of one or two methyl groups may thus be formulated as follows:

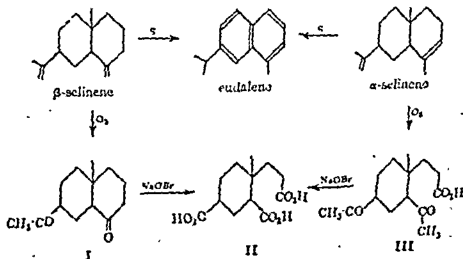


X could give two monoxides (oxidation of ring A or B), and one of these (ring B oxidised) would give VII. This, as pointed out above, was different from the monomethylcadalene actually obtained. Therefore, if X is the structure of cadinene, the monomethylcadalene obtained from cadinene must be XI. XI was synthesised, and was found to be identical with the compound obtained from cadinene. Thus X is the structure of cadinene.

It should be noted, in passing, that this new structure for cadinene has necessitated revision of the structure of copaene. Briggs and Taylor (1947), using a technique similar to that of Campbell and Soffer, have assigned the following structure to copaene.


copaene

§28a. Selinenes, $C_{15}H_{24}$. Selinene occurs in celery oil; when treated with hydrogen chloride, it forms a dihydrochloride which, when warmed with aniline, is converted into the compound $C_{15}H_{24}$. This is isomeric with selinene, and the natural compound was called *β*-selinene, and the synthetic isomer *α*-selinene (Semmler *et al.*, 1912). Semmler showed that the catalytic hydrogenation of the two selinenes gives the same tetrahydroselinene, $C_{15}H_{28}$. Thus they each contain two double bonds, and are bicyclic. Ozonolysis of *β*-selinene produces a diketone (I) with the loss of two carbon atoms, and oxidation of I with sodium hypobromite gives a



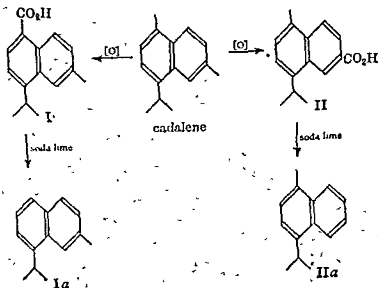
tricarboxylic acid (II), with the loss of one carbon atom. From this it follows that I contains a $CH_3 \cdot CO$— group. Ozonolysis of α-selinene gives a diketo-monocarboxylic acid (III) with loss of one carbon atom, and III, on oxidation with sodium hypobromite, loses two carbon atoms to form II. Thus III contains *two* $CH_3 \cdot CO$— groups (Semmler *et al.*, 1912). Ruzicka *et al.* (1922) distilled β-selinene with sulphur, and thereby obtained eudalene (see §28b for the evidence for the structure of this compound). If we use the isoprene rule, all the foregoing facts are explained by giving the selinenes the following structures (Ruzicka *et al.*, 1922). The relationship of the selinenes to eudesmol (§28b) confirms the nature of the carbon skeleton given to the selinenes.

§28b. **Eudesmol,** $C_{15}H_{26}O$, occurs in eucalyptus oil. Catalytic hydrogenation converts eudesmol into dihydroeudesmol, $C_{15}H_{28}O$. Thus one double bond is present in the molecule, and since eudesmol behaves as a tertiary alcohol, the parent hydrocarbon is



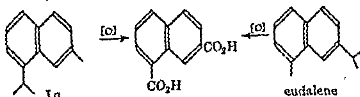
$$C_{15}H_{28} \equiv C_nH_{2n-2};$$

eudesmol is therefore bicyclic. When dehydrogenated with sulphur, eudesmol forms eudalene, $C_{14}H_{16}$, and methanethiol (Ruzicka *et al.*, 1922). Eudalene behaved as an aromatic compound (*cf.* cadalene, §28), and its structure was deduced as follows. Since eudalene was a naphthalene derivative, and since it contained one carbon atom less than cadalene, it was thought to be an *apo*-cadalene, *i.e.*, cadalene minus one methyl group. Thus eudalene is either 1-methyl-4-*iso*propylnaphthalene (II*a*) or 7-methyl-1-*iso*propylnaphthalene (I*a*). To test this hypothesis, Ruzicka oxidised cadalene with chromic acid, and thereby obtained a



[illegible] is the same as for cadalene except that ethyl malonate is used instead of ethyl methylmalonate; see §28). The synthetic compound (II*a*) was found to be different from the hydrocarbon obtained by the distillation of the naphthoic acid from cadalene. Thus the apocadalene obtained must be I*a*, *i.e.*, 7-methyl-1-*iso*propylnaphthalene.

Ruzicka now found that eudalene was not identical with either I*a* or II*a*. On oxidation, however, eudalene gives the same naphthalenedicarboxylic acid as that which is obtained by the oxidation of I*a*. This is only possible if in eudalene the two side-chains in I*a* are interchanged, *i.e.*, eudalene is 1-methyl-7-*iso*propylnaphthalene; thus:



This structure for eudalene was proved by synthesis (Ruzicka *et al.* 1922).

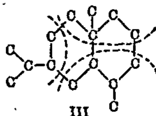
cuminal + $BrCH_2 \cdot CO_2C_2H_5$ + Zn —Reformatsky reaction→ (CHOH, CH_2, $CO_2C_2H_5$)

$-H_2O$ → (CH, CH, $CO_2C_2H_5$) —$Na-C_2H_5OH$→ (CH_2OH) —(i) HBr (ii) KCN→

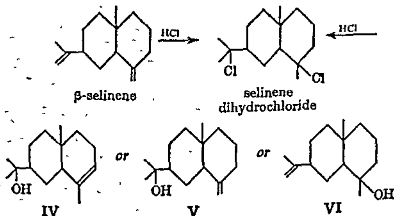
($CH_2 \cdot CN$) —(i) hydrolysis (ii) $SOCl_2$→ (COCl) —$AlCl_3$→ (O)

CH_3MgI → (HO, CH_3) —$-H_2O$→ —S→ eudalene

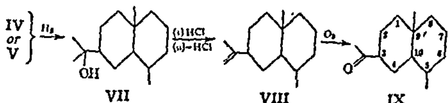
To develop the sesquiterpene carbon skeleton from that of eudalene, it is necessary to introduce one carbon atom in such a position that it is eliminated as methanethiol during the sulphur dehydrogenation (see above). If we use the *isoprene rule* with the units joined head to tail, then there is only one possible structure that fits the requirements, *viz.*, III (*cf.* §1).



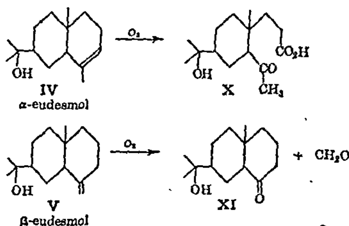
Now *β*-selinene combines with hydrogen chloride to form selinene dihydrochloride, which is also obtained by the action of hydrogen chloride on eudesmol (Ruzicka *et al.*, 1927, 1931). Since eudesmol contains one double bond and a tertiary alcoholic group, it follows that the double bond must be in the side-chain, and the hydroxyl group in the ring, or *vice versa*, *i.e.*, IV, V or VI is the structure of eudesmol.



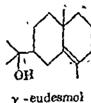
Hydrogenation of eudesmol forms dihydroeudesmol, VII, and this, on treatment with hydrogen chloride followed by boiling with aniline (to remove a molecule of hydrogen chloride), gives dihydroeudesmene, VIII. VIII, on ozonolysis, forms 3-acetyl-5.9-dimethyldecalin, IX, *with the elimination of one carbon atom*. These results are explained if IV or V is the structure of eudesmol, but not by VI. Thus the hydroxyl group is in the *iso*propyl side-chain.



The final problem is to ascertain the position of the double bond in eudesmol, *i.e.*, Is the structure IV or V? Ozonolysis of eudesmol showed that eudesmol is a mixture of IV (α-eudesmol) and V (β-eudesmol), since *two* products are obtained: a hydroxyketo-acid X, with no loss of carbon, and a hydroxyketone XI, with the loss of one carbon atom (but *cf.* citral, §5).



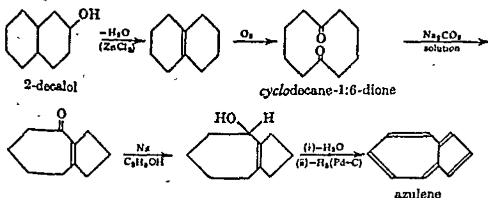
The proportions of these two isomers vary with the source, and McQuillin *et al.* (1956) have succeeded in separating them (*via* their 3:5-dinitrobenzoates), and at the same time have characterised a third, synthetic γ-isomer.



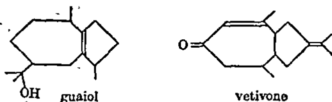
γ-eudesmol

§29. **Azulenes.** Many essential oils contain blue or violet compounds, or may form such compounds after distillation at atmo-

spheric pressure or dehydrogenation with sulphur, selenium, or palladium-charcoal (Ruzicka *et al.*, 1923). These coloured compounds may be extracted by shaking an ethereal solution of the essential oil with phosphoric acid (Sherndal, 1915). These coloured substances are known as **azulenes**. Their molecular formula is $C_{15}H_{18}$, and they are sesquiterpenes, the parent substance being azulene, $C_{10}H_8$, which contains a seven-membered ring fused to a five-membered one. Azulene has been synthesised as follows (Plattner *et al.*, 1936).



Azulene is a deep blue solid, m.p. 99°; its systematic name is bicyclo[5·3.0]decane. Two sesquiterpenes containing this bicyclodecane skeleton are



DITERPENES

§30. **Phytol**, $C_{20}H_{40}O$, b.p. 145°/0·03 mm., is an acyclic diterpene; it is produced from the hydrolysis of chlorophyll (§6. XIX), and it also forms part of the molecules of vitamins E and K (see Ch. XVII). The reactions of phytol showed that it is a primary alcohol (Willstätter *et al.*, 1907), and since on catalytic reduction phytol forms dihydrophytol, $C_{20}H_{42}O$, it therefore follows that

phytol contains one double bond. Thus the parent hydrocarbon is $C_{20}H_{42}$ ($\equiv C_nH_{2n+2}$), and so phytol is acyclic. Ozonolysis of phytol gives glycolaldehyde and a saturated ketone, $C_{18}H_{36}O$ (F. Fischer *et al.*, 1928). Thus this reaction may be written:

$$C_{16}H_{35}{=}CH\cdot CH_2OH \xrightarrow{O_3} C_{18}H_{36}O + CHO\cdot CH_2OH$$

The formula of phytol led to the suggestion that it was composed of four reduced isoprene units. If this were so, and assuming that the units are joined head to tail, the structure of the saturated ketone would be:

$$(CH_3)_2CH\cdot CH_2\cdot CH_2 \vdots CH_2\cdot CH(CH_3)\cdot CH_2\cdot CH_2 \vdots CH_2\cdot CH(CH_3)\cdot CH_2\cdot CH_2 \vdots CH_2\cdot C(CH_3){=}O$$

This structure was proved to be correct by the synthesis of the ketone from farnesol (F. Fischer *et al.*, 1928). The catalytic hydrogenation of farnesol, I, produces hexahydrofarnesol, II, which,

$$CH_3\cdot C(CH_3){=}CH\cdot CH_2\cdot CH_2\cdot C(CH_3){=}CH\cdot CH_2\cdot CH_2\cdot C(CH_3){=}CH\cdot CH_2OH \quad \text{I}$$

$$\downarrow H_2\text{–}Pd$$

$$CH_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot CH_2\cdot CH_2OH \quad \text{II}$$

$$\downarrow PBr_3$$

$$CH_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot CH_2\cdot CH_2Br \quad \text{III}$$

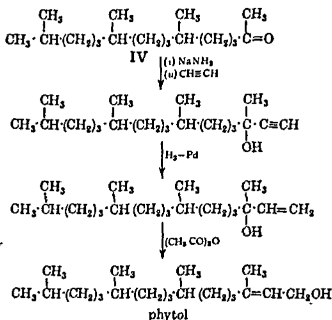
$$\downarrow CH_3\cdot CO\cdot CHNa\cdot CO_2C_2H_5$$

$$CH_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot CH_2\cdot CH_2\cdot CH(CO\cdot CH_3)(CO_2C_2H_5)$$

$$\downarrow \text{ketonic hydrolysis}$$

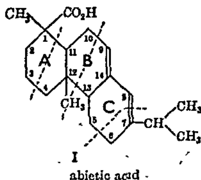
$$CH_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot(CH_2)_3\cdot CH(CH_3)\cdot CH_2\cdot CH_2\cdot CH_2\cdot CO\cdot CH_3 \quad \text{IV}$$

on treatment with phosphorus tribromide, gives hexahydrofarnesyl bromide, III. III, on treatment with sodio-acetoacetic ester, followed by ketonic hydrolysis, forms the saturated ketone, IV. This ketone (IV) was then converted into phytol as follows (F. Fischer *et al.*, 1929); it should be noted that the last step involves an allylic rearrangement (*cf.* linalool, §8).



It appears that natural phytol has a very small optical rotation; Karrer *et al.* (1943) have isolated a (+)-form from nettles.

§31. **Abietic acid**, $C_{20}H_{30}O_2$, m.p. 170–174°, is a tricyclic diterpene. The non-steam volatile residue from turpentine is known as rosin (or colophony), and consists of a mixture of resin acids which



I

abietic acid

are derived from the diterpenes. Abietic acid is one of the most useful of these acids.

A great amount of work was done before the structure of abietic acid was elucidated. For our purpose it is useful to have the structure of abietic acid as a reference, and then describe the evidence that led to this structure. I is the structure of abietic acid; the system of numbering is shown, and also the four isoprene units comprising it.

+$(CH_3)_2CHBr$ $\xrightarrow{AlCl_3}$ $(CH_3)_2CH-$ $\xrightarrow[AlCl_3]{CH_2CO\text{-}O\text{-}COCH_2}$

$(CH_3)_2CH-$ … CO, $HO_2C\cdot CH_2\cdot CH_2$ $\xrightarrow[HCl]{CH_3OH}$ $(CH_3)_2CH-$ … CO, $CH_3O_2C\cdot CH_2\cdot CH_2$ $\xrightarrow{\text{(i) } CH_3MgI \text{ (ii) heat } (-H_2O) \text{ (iii) hydrolysis}}$

$(CH_3)_2CH-$ … $C\cdot CH_3$, $HO_2C\cdot CH_2\cdot CH$ $\xrightarrow{HI-P}$ $(CH_3)_2CH-$ … $CH\,CH_3$, $HO_2C\cdot CH_2\cdot CH_2$ $\xrightarrow{H_2SO_4}$

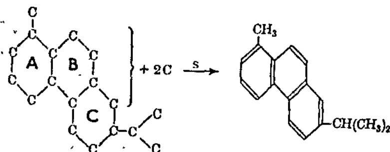
$(CH_3)_2CH-$ … CH_3, O $\xrightarrow[HCl]{Zn-Hg}$ $(CH_3)_2CH-$ … CH_3 $\xrightarrow{Se}$

$(CH_3)_2CH-$ … CH_3

retene

The general reactions of abietic acid showed that it was a monocarboxylic acid. On dehydrogenation with sulphur, abietic acid gives retene (Vesterberg, 1903); better yields of retene are obtained by dehydrogenating with selenium (Diels *et al.*, 1927), or with palladised charcoal (Ruzicka *et al.*, 1933). Retene, $C_{18}H_{18}$, m.p. 99°, was shown by oxidative degradation to be 1-methyl-7-*iso*propylphenanthrene (Bucher, 1910), and this structure was later confirmed by synthesis, *e.g.*, that of Haworth *et al.* (1932) [p. 372].

Hence we may assume that this carbon skeleton is present in abietic acid. Thus:



Now it is known that in sulphur dehydrogenations, carboxyl groups and *angular* methyl groups can be eliminated (see §2 vii. X). It is therefore possible that the two carbon atoms lost may have been originally the carboxyl group (in abietic acid) and an angular methyl group.

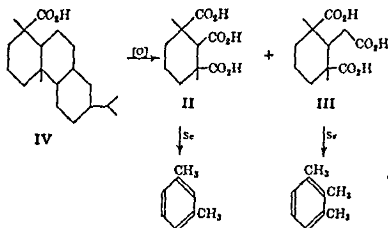
Abietic acid is very difficult to esterify, and since this is characteristic of a carboxyl group attached to a tertiary carbon atom, it suggests that abietic acid contains a carboxyl group in this state. This is supported by the fact that abietic acid evolves carbon monoxide when warmed with concentrated sulphuric acid; this reaction is also characteristic of a carboxyl group attached to a tertiary carbon atom.

Catalytic hydrogenation of abietic acid gives tetrahydroabietic acid, $C_{20}H_{34}O_2$. Thus abietic acid contains two double bonds; also, since the parent hydrocarbon is $C_{19}H_{34}$ (regarding the carboxyl group as a substituent group), abietic acid is tricyclic (parent corresponds to C_nH_{2n-4}), which agrees with the evidence already given.

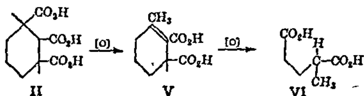
Oxidation of abietic acid with potassium permanganate gives a mixture of products, among which are two tricarboxylic acids, $C_{11}H_{16}O_6$ (II), and $C_{12}H_{18}O_6$ (III) [Ruzicka *et al.*, 1925, 1931]. II, on dehydrogenation with selenium, forms *m*-xylene, and III forms hemimellitene (1:2:3-trimethylbenzene) [Ruzicka *et al.*, 1931].

N

In both cases there is a loss of three carbon atoms, and if we assume that these were the three carboxyl groups, then two methyl groups in II and III must be in the *meta*-position. Furthermore, since II and III each contain the methyl group originally present in abietic acid (position 1), acids II and III must contain ring A of abietic acid. This suggests, therefore, that there is an angular methyl group at position 12, since it can be expect[illegible] ated from [illegible] (this 12-me[illegible] showed th[illegible] evolves two molecules of carbon monoxide when warmed with concentrated sulphuric acid; this indicates that II contains two carboxyl groups attached to tertiary carbon atoms. These results can be explained by *assuming* that one carboxyl group in II is that in abietic acid, and since in both cases this carboxyl group is attached to a tertiary carbon atom, the most likely position of this group is 1 (in abietic acid). Accepting these assumptions, the oxidation of abietic acid may be formulated as follows, also assuming IV as the carbon skeleton of abietic acid.



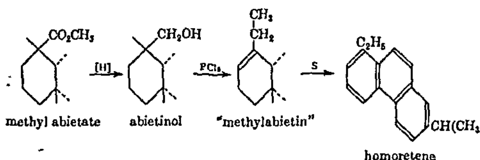
Vocke subjected II to oxidative degradation, and obtained a dicarboxylic acid (V) which, on further oxidation, gave α-methylglutaric acid (VI). Vocke assumed that II had the structure



shown, and formulated the reactions as above, *assuming* structure V as the best way of explaining the results.

Structure V (assumed by Vocke) has been confirmed by synthesis (Rydon, 1937).

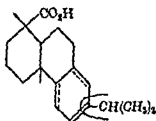
The position of the carboxyl group at position 1 in abietic acid (assumed above) has been confirmed by Ruzicka *et al.* (1922). Methyl abietate, $C_{19}H_{29}{\cdot}CO_2CH_3$, on reduction with sodium and ethanol, forms abietinol, $C_{19}H_{29}{\cdot}CH_2OH$, which, on treatment with *phosphorus pentachloride, loses a molecule of water to form* "*methylabietin*", $C_{20}H_{30}$. *This, on distillation with sulphur,* forms homoretene, $C_{19}H_{20}$. Homoretene contains one CH_2 group more than retene, and on oxidation with alkaline potassium ferricyanide, gives phenanthrene-1:7-dicarboxylic acid, the identical product obtained from the oxidation of retene under similar conditions (Ruzicka *et al.*, 1932). These results can only be explained by assuming that homoretene has an ethyl group at position 1 (instead of the methyl group in retene), *i.e.*, homoretene is 1-ethyl-7-*iso*propylphenanthrene. This has been confirmed by synthesis (Haworth *et al.*, 1932; ethylmagnesium iodide was used instead of methylmagnesium iodide in the synthesis of retene). The formation of an ethyl group in homoretene can be explained by assuming that abietinol undergoes a Wagner–Meerwein rearrangement on dehydration (see §23d). Thus:

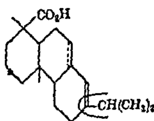


It has already been pointed out that abietic acid has two double bonds. Since abietic acid forms an adduct with maleic anhydride *at above* 100°, it was assumed that the two double bonds are conjugated (Ruzicka *et al.*, 1932). It was later shown, however, that levopimaric acid also forms the same adduct *at room temperature.* It thus appears that abietic acid isomerises to levopimaric acid at above 100°, and *then* forms the adduct. Thus this reaction cannot be accepted as evidence for conjugation in abietic acid. Nevertheless, the conjugation of the double bonds in abietic acid has been shown by means of the ultraviolet spectrum, which has not only

shown the conjugation, but also indicates that the two double bonds are *not* in the same ring (Kraft, 1935; Sandermann, 1941).

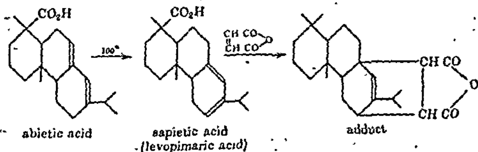
Oxidation of abietic acid with potassium permanganate gives, among other products, *iso*butyric acid (Ruzicka *et al.*, 1925). This suggests that one double bond is in ring C and the 6:7- or 7.8-position. If the double bond is in the 6:7-position, then the other double bond, which is conjugated with it, must also be in the *same* ring (5:13 or 8:14); if 7.8, then the other double bond could be in the





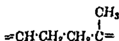
same ring C, but it could also be in ring B. Since, as we have seen, the two double bonds are in different rings, their positions are *probably* 7:8 and 14.9. Further evidence for these positions is afforded by the fact that in the oxidation of abietic acid to give acids II and III (see above), *in which ring A is intact*, rings B and C are opened, and this can be readily explained only if rings B and C each have a double bond. Oxidative studies on abietic acid by Ruzicka *et al.* (1938–1941) have conclusively confirmed the positions 7:8 and 14:9.

The only other point that will be mentioned here is the conversion of abietic acid into levopimaric acid. Since the latter was originally believed to be the enantiomorph of (+)-pimaric acid, it was called (−)-pimaric acid or lævopimaric acid. It is now known to be a *structural* isomer of dextropimaric acid, and so it has been suggested that levopimaric acid be called sapietic acid to avoid any confusion. The following equations show the formation of the adduct of abietic acid with maleic anhydride.

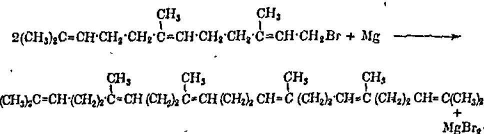


TRITERPENES

§32. **Squalene,** $C_{30}H_{50}$, b.p. 240–242°/4 mm., has been isolated from the liver oils of sharks. Catalytic hydrogenation (nickel) converts squalene into perhydrosqualene, $C_{30}H_{62}$; therefore squalene has six double bonds, and is acyclic. Ozonolysis of squalene gives, among other products, lævulic acid; this suggests that the following group is present in squalene:



Since squalene cannot be reduced by sodium and amyl alcohol, there are no conjugated double bonds present in the molecule. Perhydrosqualene was found to be identical with the product obtained by subjecting hexahydrofarnesyl bromide to the Wurtz reaction. This led Karrer *et al.* (1931) to synthesise squalene itself from farnesyl bromide by a Wurtz reaction.



It should be noted that the centre portion of the squalene molecule has the two isoprene units joined tail to tail (*cf.* the carotenoids, Ch. IX). Squalene forms a thiourea inclusion complex, and hence it has been inferred that it is the all-*trans* stereoisomer (Schiessler *et al.*, 1952).

§32a. **Biosynthesis of terpenes.** As more and more natural products were synthesised in the laboratory, so grew the interest in how these compounds are synthesised in the living organism (both animal and plant). The general approach to biosynthesis has been to break up the structure into units from which the compound could plausibly be derived. These units must, however, be known, or can be expected, to be available in the organism. Furthermore, this does not mean that the units chosen must necessarily be involved in the building-up of the compound. The general principle is that although a particular unit may itself be involved,

it is also possible that its "equivalent" may act as a substitute, *i.e.*, any compound that can readily give rise to this unit (by means of various reactions such as reduction, oxidation, etc.) may be the actual compound involved in the biosynthesis. *E.g.*, the equivalent of formaldehyde could be formic acid, and that of acetone acetoacetic acid. One other point about the choice of units or their equivalents is to attempt to find some relationships between the various groups of natural products so that the units chosen are *common* precursors.

When the units have been chosen, the next problem is to consider the types of reactions whereby the natural products are synthesised in the organism. The general principle is to use reactions which have been developed in the laboratory. The difficulty here is that some types of laboratory reactions require conditions that cannot operate in the organism, *e.g.*, carboxylation and decarboxylation are known biological processes, but when carried out in the laboratory, these reactions normally require elevated temperatures. Deamination is also a known biological process, but in the laboratory this reaction is usually carried out under conditions (of *p*H) which would be lethal to the living organism. These differences between laboratory syntheses and biosyntheses are due to the action of enzymes in the latter. According to Schöpf (1932), syntheses in plants may take place through the agency of specific or non-specific enzymes (see §§12–17. XIII), or without enzymes at all. Chemical syntheses (these do not involve the use of enzymes) must therefore, from the point of biosynthetic studies, be carried out under conditions of *p*H and temperatures comparable with those operating in plants. Chemical syntheses performed in this way (with the suitable units) are said to be carried out under *physiological conditions* (which involve a *p*H of about 7 in aqueous media and ordinary temperatures).

Reactions which are commonly postulated in biosynthesis are oxidation, hydrogenation, dehydrogenation, dehydration, esterification, hydrolysis, carboxylation, decarboxylation, amination, deamination, isomerisation, condensation, and polymerisation. It might be noted here that the choice of units and type of reaction are usually dependent on each other. Furthermore, other reactions which are known to occur in biological syntheses are *O*- and *N*-methylation or acylation. These may be described as *extra-skeletal processes*, and can occur at any suitable stage in the postulated biosynthesis. Another extra-skeletal process is *C*-methylation, but this is much rarer than those mentioned above.

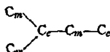
Now let us apply these principles to the biosynthesis of terpenes. As we have seen, according to the special isoprene rule, terpenes

are built up of isoprene units joined head to tail (§1). Assuming then that the isoprene unit is the basic unit, the problem is: How is it formed, and how do these units join to form the various types of terpenes? At present it is believed that the fundamental units used in the cell in syntheses are water, carbon dioxide, formic acid (as "active formate"), and acetic acid (as "active acetate"). These "active" compounds are acyl derivatives of coenzyme A (written as CoA—H in the following equation); *e.g.*, acetoacetic acid is believed to be formed as follows:



$$2CH_3{\cdot}CO{\cdot}CoA + H_2O \rightarrow CH_3{\cdot}CO{\cdot}CH_2{\cdot}CO_2H + 2CoA{-}H$$

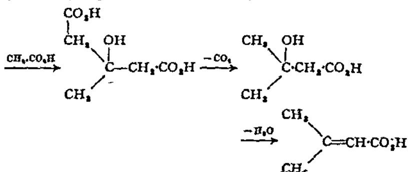
Now the biosynthesis of cholesterol (§7a. XI) from acetic acid labelled with ^{14}C in the methyl group (C_m) and in the carboxyl group (C_c) has led to the suggestion that the carbon atoms in the isoprene unit are distributed as follows:



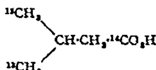
This distribution is in agreement with the following scheme in which senecioic acid (3-methylbut-2-enoic acid) is formed first.



$$CH_3{\cdot}CO_2H + CH_3{\cdot}CO_2H \rightarrow CH_3{\cdot}CO{\cdot}CH_2{\cdot}CO_2H$$

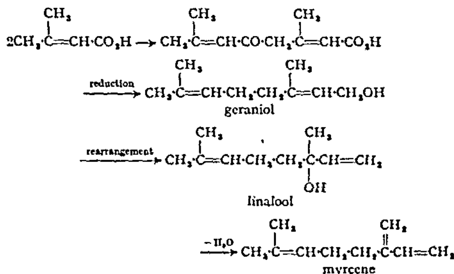


Support for the formation of this carbon skeleton is given by the fact that labelled *iso*valeric acid gives rise to cholesterol in which the *iso*propyl group and the carboxyl group have been incorporated.



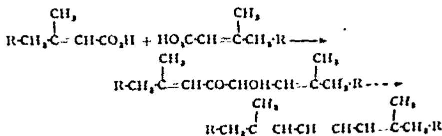
Senecioic acid occurs naturally (in certain species of *Senecio*), and condensations between two or more molecules can now occur

as with acetic acid to produce acyclic terpene skeletons (the feasibility of these condensations is in accord with the principle of vinology; see Vol. I). By then postulating various processes at suitable stages (the order of which is not certain), various acyclic terpenes are produced, *e g*,



Cyclisation of suitable intermediates would produce monocyclic terpenes, *e g*., geraniol would give α-terpineol (see §7).

The condensations discussed so far lead to head to tail unions of the isoprene units. As we have seen, some unions (in the centre of the molecule) are tail to tail, *e g*., squalene is composed of two farnesyl chains joined end to end (§32). It has been suggested that such tail to tail unions are produced by the acyloin type of condensation (R is some radical composed of isoprene units joined head to tail):



POLYTERPENES

§33. **Rubber.** *Rubber* (*caoutchouc*) is obtained from latex, which is an emulsion of rubber particles in water that is obtained from the inner bark of many types of trees which grow in the tropics and sub-tropics. When the bark of the rubber tree is cut, latex slowly exudes from the cut. Addition of acetic acid coagulates the rubber, which is then separated from the liquor and either pressed into blocks or rolled into sheets, and finally dried in a current of warm air, or smoked.

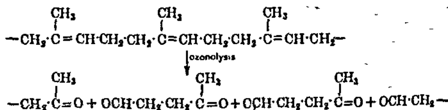
Crude latex rubber contains, in addition to the actual rubber hydrocarbons (90–95 per cent.), proteins, sugars, fatty acids and resins, the amounts of these substances depending on the source. Crude rubber *is soft and sticky, becoming more so as the temperature* rises. It has a low tensile strength and its elasticity is exhibited only over a narrow range of temperature. Crude rubber dissolves in many organic solvents, *e.g.*, benzene, ether, light petrol; it also swells when it comes into contact with these solvents. On the other hand, rubber is insoluble in acetone, methanol, etc. When unstretched, rubber is amorphous; stretching or prolonged cooling causes rubber to crystallise.

Structure of rubber. The destructive distillation of rubber gives isoprene as one of the main products; this led to the suggestion that rubber is a polymer of isoprene, and therefore to the molecular formula $(C_5H_8)_n$. This molecular formula has been confirmed by the analysis of pure rubber. Crude rubber may be purified by fractional precipitation from benzene solution by the addition of acetone. This fractional precipitation, however, produces molecules of different sizes, as shown by the determination of the molecular weights of the various fractions by osmotic pressure, viscosity and ultracentrifuge measurements; molecular weights of the order of 300,000 have been obtained.

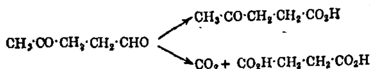
The halogens and the halogen acids readily add on to rubber, *e.g.*, bromine gives an addition product of formula $(C_5H_8Br_2)_n$, and hydrogen chloride the addition product $(C_5H_9Cl)_n$. Pure rubber has been hydrogenated to the fully saturated hydrocarbon $(C_5H_{10})_n$ —this is known as *hydrorubber*—by heating with hydrogen in the presence of platinum as catalyst (Pummerer *et al.*, 1922). Rubber also forms an ozonide of formula $(C_5H_8O_3)_n$. All these addition reactions clearly indicate that rubber is an unsaturated compound, and the formulae of the addition products show that there is one double bond for each isoprene unit present.

Ozonolysis of rubber produces *lævulaldehyde* and its peroxide,

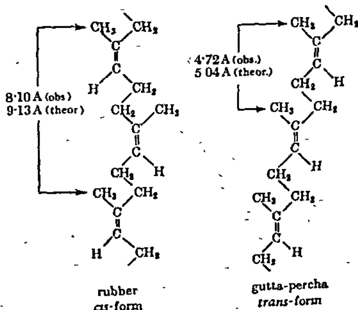
lævulic acid and small amounts of carbon dioxide, formic acid and succinic acid (Harries, 1905–1912). Pummerer (1931) showed that the lævulic derivatives comprised about 90 per cent. of the products formed by the ozonolysis. This observation led to the suggestion that rubber is composed of isoprene units joined head to tail. Thus, if rubber has the following structure, the formation of the products of ozonolysis can be explained:



Some of the lævulaldehyde is further oxidised to lævulic and succinic acids.



Gutta-percha (which is also obtained from the bark of various trees) is isomeric with rubber; their structures are the same, as shown by the methods of analysis that were used for rubber.



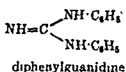
X-ray diffraction studies (Bunn, 1942) have shown that rubber is composed of long chains built up of isoprene units arranged in the *cis*-form, whereas gutta-percha is the *trans*-form.

In rubber, the chain repeat unit is 8·10 A, whereas in gutta-percha it is 4·72 A. Both of these values are shorter than the theoretical values of the repeat distances (9·13 A and 5·04 A respectively) calculated from models. The reasons for these discrepancies are not clear, but for gutta-percha it has been explained by assuming that the isoprene units are not coplanar. The infra-red absorption spectrum of rubber has bands which are in keeping with the structure that has been proposed. Also, the *linear* shape of the molecule is indicated by viscosity measurements of rubber solutions.

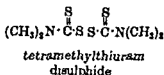
§33a. Vulcanisation of rubber. When crude rubber is heated with a few per cent. of sulphur, the rubber becomes *vulcanised*. Vulcanised rubber is less sticky than crude rubber, and is not so soluble and does not swell so much in organic solvents. Furthermore, vulcanised rubber has greater tensile strength and elasticity than crude rubber.

The mechanism of vulcanisation is still not clear. Vulcanised rubber is not so unsaturated as rubber itself, the loss of one double bond corresponding approximately to each sulphur atom introduced. It therefore appears that *some* sulphur atoms enter the chain, vulcanisation thus occurring through intramolecular and intermolecular cross-links; it is the latter type of reaction that is desirable in vulcanisation. It should be noted that not all the sulphur is in a combined state; some is *free*, and this can be readily extracted.

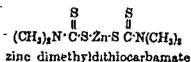
Vulcanisation may be accelerated and carried out at lower temperatures in the presence of certain organic compounds. These compounds are consequently known as *accelerators*, and all of them contain nitrogen or sulphur, or both, *e.g.*,



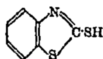
diphenylguanidine



tetramethylthiuram disulphide



zinc dimethyldithiocarbamate



mercaptobenzothiazole

Mercaptobenzothiazole is the most widely used accelerator. Many

inorganic compounds can also act as accelerators, *e.g.*, zinc oxide. Organic accelerators are promoted by these inorganic compounds, and current practice is to vulcanise rubber with, *e.g.* mercapto-

of sulphur used, the best physical properties apparently being achieved by using about 3 per cent. sulphur, 5 per cent. zinc oxide, and about 1 per cent. of the accelerator. When 30–50 per cent. sulphur is used, the product is *ebonite*.

The elasticity of rubber is believed to be due to the existence of rubber as long chain molecules which are highly "kinked" in the normal state. When subjected to a stretching force, these chains "unkink", and return to their normal condition when the force is removed.

§33b. Synthetic rubbers. There are many synthetic rubbers in use, each type possessing certain desirable properties.

Buna rubbers. Under the influence of sodium, butadiene polymerises to a substance which has been used as a rubber substitute under the name of *Buna* (see Vol. I). *Buna N* is a synthetic rubber which is produced by the copolymerisation of butadiene and vinyl cyanide. *Buna S* or *Perbunan* is a copolymer of butadiene and styrene.

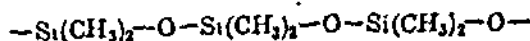
Butyl rubber. Copolymerisation of *iso*butylene with a small amount of isoprene produces a poly*iso*butylene known as *Butyl rubber*.

Neoprene. When passed into a solution of cuprous chloride in ammonium chloride, acetylene dimerises to vinylacetylene. This dimer can add on one molecule of hydrogen chloride to form *Chloroprene* (2-chlorobuta-1·3-diene), the addition taking place in accordance with Markownikoff's rule (see also Vol. I).



Chloroprene readily polymerises to a rubber-like substance known as *Neoprene*. Actually, the nature of the polychloroprene depends on the conditions of the polymerisation.

Silicone rubbers. These are chemically similar to the silicone resins. The chief silicone rubber is prepared by treating the hydrolysis product of dimethyldichlorosilane, $(CH_3)_2SiCl_2$, with various compounds capable of increasing the molecular weight without the formation of cross-links, *i.e.*, they produce long-chain molecules.



inorganic compounds can also act as accelerator Organic accelerators are promoted by these inor and current practice is to vulcanise rubber wit benzothiazole in the presence of zinc oxide.

The actual properties of vulcanised rubber der of sulphur used, the best physical propertie achieved by using about 3 per cent. sulphur, 5 p and about 1 per cent. of the accelerator. W sulphur is used, the product is *ebonite*.

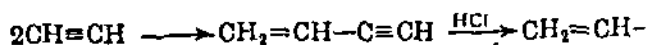
The elasticity of rubber is believed to be du rubber as long chain molecules which are hig normal state. When subjected to a stretchin " unkink ", and return to their normal condit removed

§33b. Synthetic rubbers. There are ma in use, each type possessing certain desirab

Buna rubbers. Under the influence of polymerises to a substance which has been use tute under the name of *Buna* (see Vol. I). B rubber which is produced by the copolymerisat vinyl cyanide. *Buna S* or *Perbunan* is a cop and styrene.

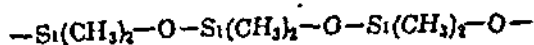
Butyl rubber. Copolymerisation of *iso*but amount of isoprene produces a poly*iso*butylen *rubber.*

Neoprene. When passed into a solution o in ammonium chloride, acetylene dimerises to vin dimer can add on one molecule of hydrogen *Chloroprene* (2-chlorobuta-1:3-diene), the addition accordance with Markownikoff's rule (see also Vo


$$2CH{\equiv}CH \longrightarrow CH_2{=}CH{-}C{\equiv}CH \xrightarrow{HCl} CH_2{=}CH{-}$$

Chloroprene readily polymerises to a rubber-like sub as *Neoprene* Actually, the nature of the polychloro on the conditions of the polymerisation.

Silicone rubbers. These are chemically similar to resins. The chief silicone rubber is prepared by hydrolysis product of dimethyldichlorosilane, $(CH_3)_2$ various compounds capable of increasing the molecu without the formation of cross-links, *i.e.*, they produce molecules.


$$-Si(CH_3)_2-O-Si(CH_3)_2-O-Si(CH_3)_2-O-$$

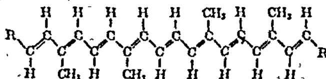
CHAPTER IX

CAROTENOIDS

§1. Introduction. The carotenoids are yellow or orange pigments which are widely distributed in plants and animals. The amounts in which they occur are very small, and it is not certain whether animals can actually synthesise carotenoids. Chlorophyll is always associated with the carotenoids carotene and lutein; the carotenoids act as photosensitisers in conjunction with chlorophyll. When chlorophyll is absent, *e.g.*, in fungi, then the carotenoids are mainly responsible for colour. Carotenoids are also known as lipochromes or chromolipids because they are fat-soluble pigments. They give a deep blue colour with concentrated sulphuric acid and with a chloroform solution of antimony trichloride (the Carr-Price reaction); this Carr-Price reaction is the basis of one method of the quantitative estimation of carotenoids. Some carotenoids are hydrocarbons; these are known as the *carotenes*. Other carotenoids are oxygenated derivatives of the carotenes; these are the *xanthophylls*. There are also acids, the *carotenoid acids*, and esters, the *xanthophyll esters*.

Chemically, the carotenoids are polyenes, and almost all the carotenoid hydrocarbons have the molecular formula $C_{40}H_{56}$. Also, since the carbon skeleton of these compounds has a polyisoprene structure, they may be regarded as tetraterpenes (*cf.* §1. VIII).

In most of the carotenoids, the central portion of the molecule is composed of a long conjugated chain comprised of four isoprene units, the centre two of which are joined tail to tail. The ends of the chain may be two open-chain structures, or one open-chain structure and one ring, or two rings. The colour of the carotenoids is attributed to the extended conjugation of the central chain (see Vol. I). X-ray analysis has shown that in the majority of natural carotenoids, the double bonds are in the *trans* position; a few natural carotenoids are *cis*. Thus, if we represent the ends of the chain by R (where R may be an open-chain structure or a ring system), *trans*-carotenes may be written:



Whitby and Katz, Synthetic Rubber, *Ind. Eng. Chem.*, 1933, **25**, 1204, 1338.

Midgley, Critical Examination of Some Concepts in Rubber Chemistry, *Ind Eng Chem*, 1942, **34**, 891.

Flory, *Principles of Polymer Chemistry*, Cornell University Press (1953).

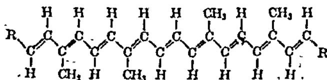
CHAPTER IX

CAROTENOIDS

§1. Introduction. The carotenoids are yellow or orange pigments which are widely distributed in plants and animals. The amounts in which they occur are very small, and it is not certain whether animals can actually synthesise carotenoids. Chlorophyll is always associated with the carotenoids carotene and lutein; the carotenoids act as photosensitisers in conjunction with chlorophyll. When chlorophyll is absent, *e g.*, in fungi, then the carotenoids are mainly responsible for colour. Carotenoids are also known as lipochromes or chromolipids because they are fat-soluble pigments. They give a deep blue colour with concentrated sulphuric acid and with a chloroform solution of antimony trichloride (the Carr-Price reaction); this Carr-Price reaction is the basis of one method of the quantitative estimation of carotenoids. Some carotenoids are hydrocarbons; these are known as the *carotenes*. Other carotenoids are oxygenated derivatives of the carotenes; these are the *xanthophylls*. There are also acids, the *carotenoid acids*, and esters, the *xanthophyll esters*.

Chemically, the carotenoids are polyenes, and almost all the carotenoid hydrocarbons have the molecular formula $C_{40}H_{56}$. Also, since the carbon skeleton of these compounds has a polyisoprene structure, they may be regarded as tetraterpenes (*cf.* §1. VIII).

In most of the carotenoids, the central portion of the molecule is composed of a long conjugated chain comprised of four isoprene units, the centre two of which are joined tail to tail. The ends of the chain may be two open-chain structures, or one open-chain structure and one ring, or two rings. The colour of the carotenoids is attributed to the extended conjugation of the central chain (see Vol. I). X-ray analysis has shown that in the majority of natural carotenoids, the double bonds are in the *trans* position; a few natural carotenoids are *cis*. Thus, if we represent the ends of the chain by R (where R may be an open-chain structure or a ring system), *trans*-carotenes may be written:



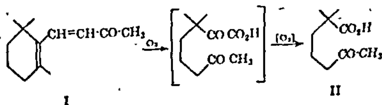
§2. Carotenes. Carotene was first isolated by Wackenroder (1831) from carrots (this was the origin of the name *carotin*, which was later changed to *carotene*). The molecular formula of carotene, however, was not determined until 1907, when Willstätter showed it was $C_{40}H_{56}$. Carotene was shown to be unsaturated, and when treated with a *small* amount of iodine, it forms a crystalline di-iodide, $C_{40}H_{56}I_2$. Kuhn (1929) separated this di-iodide into two fractions by means of fractional crystallisation. Treatment of each fraction with thiosulphate regenerated the corresponding carotenes, which were designated α- and β-carotene. Kuhn *et al.* (1933) then found that chromatography gives a much better separation of the carotenes themselves, and in this way isolated a third isomer, which he designated γ-carotene.

α-Carotene, m.p. 187–187·5°; optically active (dextrorotatory).
β-Carotene, m.p. 184·5°; optically inactive.
γ-Carotene, m.p. 176·5°; optically inactive.

It appears that all three carotenes occur together in nature, but their relative proportions vary with the source, *e.g.*, carrots contain 15 per cent. α, 85 per cent. β, and 0·1 per cent. γ. Carotenes are obtained commercially by chromatography, two of the best sources being carrots and alfalfa.

§3. β-Carotene, $C_{40}H_{56}$. When catalytically hydrogenated (platinum), β-carotene forms perhydro-β-carotene, $C_{40}H_{78}$. Thus β-carotene contains eleven double bonds, and since the formula of perhydro-β-carotene corresponds to the general formula C_nH_{2n-2}, it follows that the compound contains two rings.

When exposed to air, β-carotene develops the odour of violets. Since this odour is characteristic of β-ionone, it was thought that this residue is present in β-carotene (see §6. VIII). This was confirmed by the fact that the oxidation of a benzene solution of β-carotene with cold aqueous potassium permanganate gives β-ionone. Now β-ionone, I, on ozonolysis, gives among other things, geronic acid, II (Karrer *et al.*, 1929).



β-Carotene, on ozonolysis, gives geronic acid in an amount that

corresponds to the presence of two β-ionone residues (Karrer *et al.*, 1930). Thus a tentative structure for β-carotene is:

$$\text{C}_{13}\ (\beta\text{-ionone ring})\text{—CH=CH·C(CH}_3\text{)=}\ [\longleftarrow \text{C}_{14} \longrightarrow]\ \text{=C(CH}_3\text{)·CH=CH—}(\beta\text{-ionone ring})\ \text{C}_{13}$$

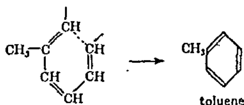
Since the colour of β-carotene is due to extended conjugation (§1), the C_{14} portion of the molecule will be conjugated. The presence of conjugation in this central portion is confirmed by the fact that β-carotene forms an adduct with five molecules of maleic anhydride (Nakamiya, 1936).

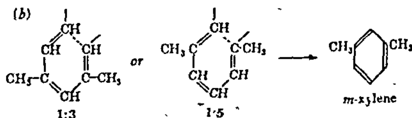
Geronic acid, on oxidation with cold aqueous potassium permanganate, forms a mixture of acetic acid, α:α-dimethylglutaric, III, α:α-dimethylsuccinic, IV, and dimethylmalonic acids, V.

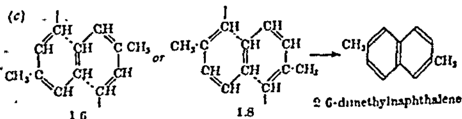
$$\underset{\text{II}}{\text{geronic acid}} \xrightarrow{[O]} CH_3·CO_2H + \underset{\text{III}}{HO_2C·C(CH_3)_2·CH_2·CH_2·CO_2H} \xrightarrow{[O]} \underset{\text{IV}}{HO_2C·C(CH_3)_2·CH_2·CO_2H} \xrightarrow{[O]} \underset{\text{V}}{HO_2C·C(CH_3)_2·CO_2H}$$

Oxidation of β-carotene in benzene solution with cold aqueous permanganate gives a mixture of β-ionone, III, IV, V, and acetic acid, the amount of acetic acid being more than can be accounted for by the presence of two β-ionone residues. Thus there must be some methyl side-chains in the central C_{14} portion of the molecule. Since it is essential to know the exact number of these methyl side-chains, this led to the development of the **Kuhn-Roth methyl side-chain determination** (1931). The first method used was to oxidise the carotenoid with alkaline permanganate, but later chromic acid (chromium trioxide in sulphuric acid) was found to be more reliable, the methyl group in the fragment $—C(CH_3)=$ being always oxidised to acetic acid. It was found that alkaline permanganate only oxidises the fragment $=C(CH_3)—CH=$ to acetic acid, and fragments such as $=C(CH_3)—CH_2—$ are incompletely oxidised to acetic acid, or not attacked at all (Karrer *et al.*, 1930). Since a molecule ending in an *iso*propylidene group also gives acetic acid on oxidation with chromic acid, this end group is determined by ozonolysis, the acetone so formed being estimated volumetrically. Application of the Kuhn-Roth methyl side-chain determination to

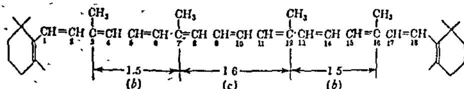
β-carotene gave four molecules of acetic acid, thus indicating that there are four $-C(CH_3)=$ groups in the *chain*. The positions of two of these have already been tentatively placed in the two end β-ionone residues (see tentative structure above), and so the problem is now to find the positions of the remaining two. This was done as follows. Distillation of carotenoids under normal conditions brings about decomposition with the formation of aromatic compounds. Thus the distillation of β-carotene produces toluene, *m*-xylene and 2,6-dimethylnaphthalene (Kuhn *et al.*, 1933). The formation of these compounds may be explained by the cyclisation of fragments of the polyene chain, without the β-ionone rings being involved. The following types of chain fragments would give the desired aromatic products:



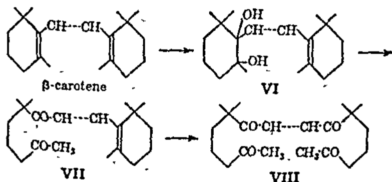




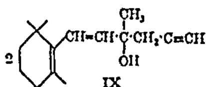
The following *symmetrical* structure for β-carotene would satisfy the requirements of (*a*), (*b*) and (*c*); the tail to tail union of the two isoprene units at the centre should be noted.



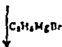
This symmetrical formula for β-carotene has been confirmed by the following oxidation experiments (Kuhn *et al.*, 1932–1935). When β-carotene is oxidised *rapidly* with potassium dichromate, dihydroxy-β-carotene, VI, is obtained and this, on oxidation with lead tetra-acetate, gives semi-β-carotenone, VII, a diketone. Since both VI and VII contain the *same* number of carbon atoms as β-carotene, it follows that the *double bond in one of the β-ionone rings* has been oxidised; otherwise there would have been chain scission had the chain been oxidised. Oxidation of semi-β-carotenone with chromium trioxide produces β-carotenone, VIII, a tetraketone which also has the same number of carbon atoms as β-carotene. Thus, in this compound, the *other* β-ionone ring is opened. Now only *one* dihydroxy-β-carotene and *one* semi-β-carotenone are obtained, and this can be explained only by assuming a symmetrical structure for β-carotene. Thus the oxidations may be formulated:

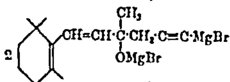


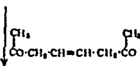
This structure for β-carotene has been confirmed by synthesis, *e.g.*, that of Karrer *et al.* (1950). The acetylenic carbinol IX is treated with ethylmagnesium bromide, and the product is treated as shown.

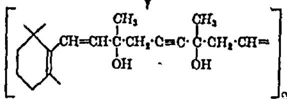


IX

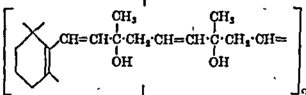


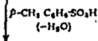


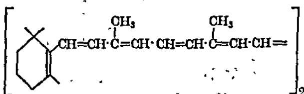






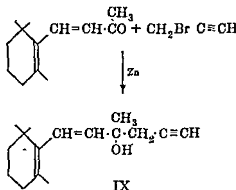




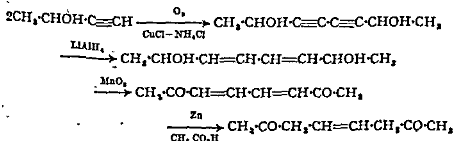


β-carotene

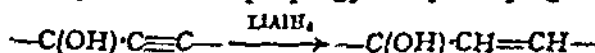
IX has been prepared by Isler (1949) by treating β-ionone with propargyl bromide in the presence of zinc (*cf.* the Reformatsky reaction):



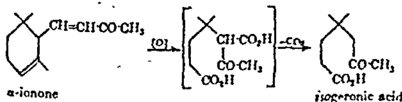
The most convenient way of preparing the diketone (oct-4-ene-2:7-dione) starts with but-1-yn-3-ol (Inhoffen *et al.*, 1951):



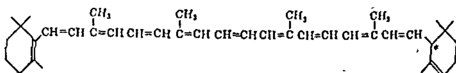
An important point to note in this synthesis is that lithium aluminium hydride will reduce a triple bond to a double bond when the former is adjacent to a propargylic hydroxyl group, *i.e.*,



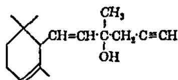
§4. **α-Carotene, $C_{40}H_{56}$.** This is isomeric with β-carotene, and oxidation experiments on α-carotene have led to results similar to those obtained for β-carotene, except that *iso*geronic acid is obtained as well as geronic acid. Since *iso*geronic acid is an oxidation product of α-ionone, the conclusion is that α-carotene contains one β-ionone ring and one α-ionone ring (§6. VIII) [Karrer *et al.*, 1933].



Thus the structure of α-carotene is:



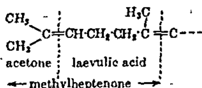
As we have seen, α-carotene is optically active (§1), and this is due to the presence of the asymmetric carbon atom (*) in the α-ionone ring. The structure given for α-carotene has been confirmed by synthesis (Karrer *et al.*, 1950). The method is the same as that described for β-carotene, except that *one* molecule of the acetylenic alcohol (structure IX, §3) is used together with one molecule of the corresponding α-ionone derivative:



It is interesting to note that α-carotene has been converted into the β-isomer by heating the α-compound with ethanolic sodium ethoxide and benzene at 100–110° for some time (Karrer *et al.*, 1947); this is an example of *three carbon prototropy*.

§5. **Lycopene,** $C_{40}H_{56}$, m.p. 175°, is a carotenoid that is the tomato pigment. *Since the structure of γ-carotene depends on that of lycopene, the latter will be discussed here, and the former in the next section.*

On catalytic hydrogenation (platinum), lycopene is converted into perhydrolycopene, $C_{40}H_{82}$. Therefore lycopene has thirteen double bonds, and is an acyclic compound (Karrer *et al.*, 1928). Ozonolysis of lycopene gives, among other products, acetone and lævulic acid; this suggests that lycopene contains the terminal residue:



```
                          CH3            CH3                 CH3
                          |              |                   |
(CH3)2C=CH CH2 CH2·C=CH·CH=CH·C=CH·CH=CH·C=CH·CH
                                                              ||
(CH3)2C=CH·CH2·CH2·C=CH· CH=CH·C=CH·CH=CH·C=CH·CH
                          |              |                   |
                          CH3            CH3                 CH3
```

lycopene

↓ CrO_3

```
                     CH3                       CH3          CH3
                     |                         |            |
(CH3)2C=CH·CH2·CH2·C=O + CHO·CH=CH·C=CH·CH=CH·C=CH·CH
methylheptenone                                                ||
   (CH3)2C=CH·CH2·CH2·C=CH·CH=CH C=CH·CH=CH·C=CH·CH
                       |              |             |
                       CH3            CH3           CH3
```

lycopenal

↓ CrO_3

```
                                          CH3           CH3
                                          |             |
                  CH3     CHO·CH=CH·C=CH·CH=CH·C=CH·CH
                  |    +                                  ||
(CH3)2C=CH·CH2·CH2·C=O    CHO·CH=CH·C=CH·CH=CH·C=CH·CH
                                          |             |
methylheptenone                           CH3           CH3
```

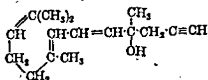
bixindialdehyde

↓ (i) NH_2OH (ii) $(CH_3CO)_2O$ [$-H_2O$] (iii) hydrolysis

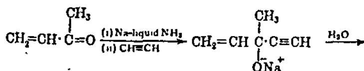
```
                      CH3           CH3
                      |             |
CO2H·CH=CH·C=CH·CH=CH·C=CH CH
                                       ||
CO2H·CH=CH·C=CH·CH=CH·C=CH CH
                      |             |
                      CH3           CH3
```

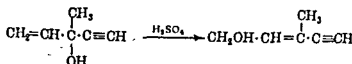
norbixin

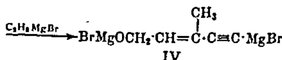
The structure assigned to lycopene has been confirmed by synthesis (Karrer *et al.*, 1950). Instead of the acetylenic carbinol IX in §3, two molecules of the following compound were used.



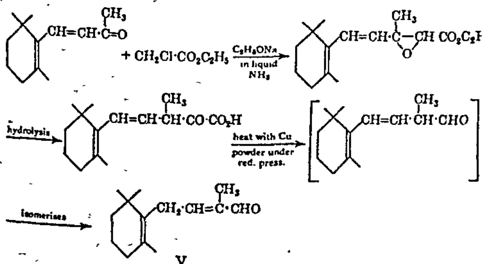
Preparation of IV.







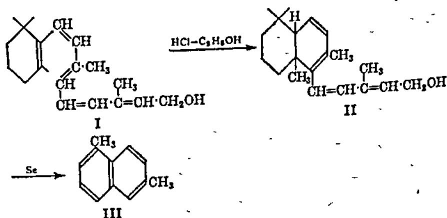
Preparation of V.



In the hydrogenation of VI to VII, barium sulphate is used to act as a poison to the catalyst to prevent hydrogenation of the *double* bonds. Partial acetylation of VII (primary alcoholic groups are more readily acetylated than secondary) protects the terminal group from an allylic rearrangement in the conversion of VIII to IX.

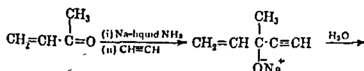
The crude vitamin A_1 obtained in the above synthesis was purified *via* its ester with anthraquinone-2-carboxylic acid, and was thereby obtained in a crystalline form which was shown to be identical with natural vitamin A_1.

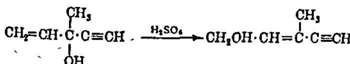
bonds. Since vitamin A_1 forms an ester with *p*-nitrobenzoic acid (this ester is not crystallisable), it follows that vitamin A_1 contains a hydroxyl group. Thus the parent hydrocarbon of vitamin A_1 is $C_{20}H_{40}$, and consequently the molecule contains one ring. Ozonolysis of vitamin A_1 produces one molecule of geronic acid (§3) per molecule of vitamin A_1, and so there must be one β-ionone nucleus present (Karrer, 1931, 1932). Oxidation of vitamin A_1 with permanganate produces acetic acid; this suggests that there are some $—C(CH_3)=$ groups in the chain. All of the foregoing facts are in keeping with the suggestion that vitamin A_1 is half the β-carotene structure. When heated with an ethanolic solution of hydrogen chloride, vitamin A_1 is converted into some compound (II) which, on dehydrogenation with selenium, forms 1:6-dimethylnaphthalene, III (Heilbron *et al.*, 1932). Heilbron *assumed* I as the structure of vitamin A_1, and explained the course of the reaction as follows:

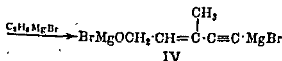


Perhydrovitamin A_1 has been synthesised from β-ionone (Karrer, 1933), and was shown to be identical with the compound obtained by reducing vitamin A_1; thus there is evidence to support the structure assigned to vitamin A_1. Final proof of structure must rest with a synthesis of vitamin A_1 itself, and this has now been accomplished by several groups of workers. The following synthesis is that of Isler *et al.* (1947). This starts with methyl vinyl ketone to produce compound IV, one stage of the reactions involving an allylic rearrangement (*cf.* §8. VIII). Compound V is prepared from β-ionone by means of the Darzens glycidic ester reaction (see also Vol. I). The following chart shows the steps of the synthesis, and it should be noted that another allylic rearrangement is involved in one of the later steps (see p. 400):

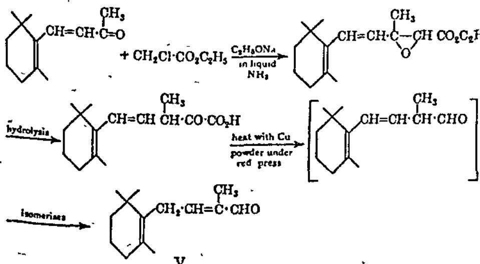
Preparation of IV.







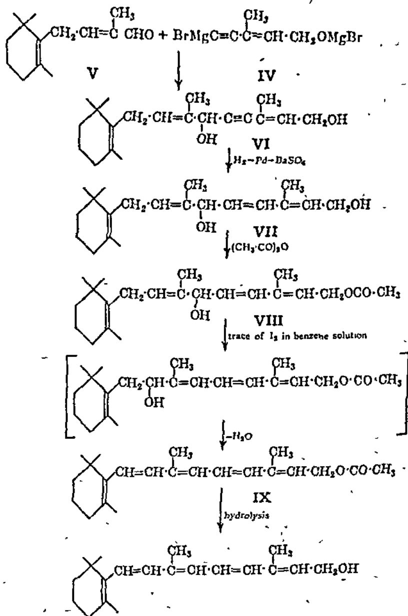
Preparation of V.



In the hydrogenation of VI to VII, barium sulphate is used to act as a poison to the catalyst to prevent hydrogenation of the *double* bonds. Partial acetylation of VII (primary alcoholic groups are more readily acetylated than secondary) protects the terminal group from an allylic rearrangement in the conversion of VIII to IX.

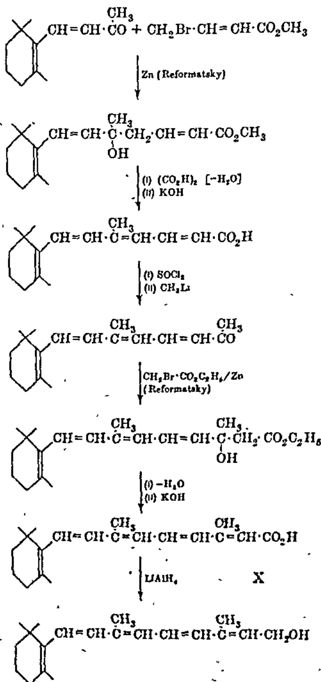
The crude vitamin A_1 obtained in the above synthesis was purified *via* its ester with anthraquinone-2-carboxylic acid, and was thereby obtained in a crystalline form which was shown to be identical with natural vitamin A_1.

Combination of IV and V, etc.



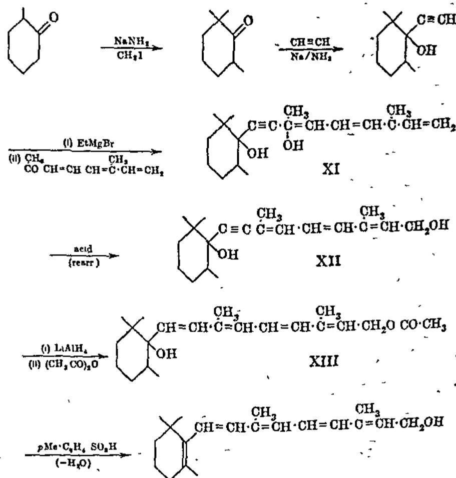
Lindlar (1952) has shown that triple bonds may be partially
... in the presence of a *Pd—CaCO$_3$ catalyst that has*
...d acetate; better
Thus the hydro-

genation of VI gives VII in 86 per cent. yield when the Lindlar catalyst is used.



Another method of synthesising vitamin A_1 is due to van Dorp *et al.* (1946) who prepared vitamin A_1 acid (X), which was then reduced by means of lithium aluminium hydride to vitamin A_1 by Tishler (1949); β-ionone and methyl γ-bromocrotonate are the starting materials (see p. 401).

Attenburrow *et al.* (1952) have also synthesised vitamin A_1 starting from 2-methyl*cyclo*hexanone.



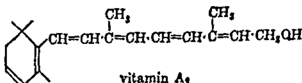
Acid causes rearrangement of XI to XII in which all multiple bonds are in complete conjugation, and the reduction of XII to XIII by lithium aluminium hydride is possible because of the presence of the propargylic hydroxyl grouping (§3).

Synthetic vitamin A_1 is now a commercial product.

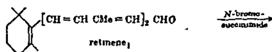
Neovitamin A_1. An isomeric biologically active form of vitamin A_1 has been isolated from fish-liver oils (Robeson *et al.*, 1947). It

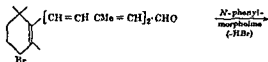
has a m.p. 59–60°, and it appears that this isomer, neovitamin A_1, is a geometrical isomer of vitamin A_1, the former being the 2-*cis*-form, and the latter the all-*trans*-form, the double bond involved in the isomerism being the terminal one, *i.e.*, the double bond in the grouping $—C(CH_3)=CH·CH_2OH$.

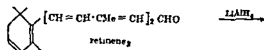
Vitamin A_2. A second vitamin A, vitamin A_2, has been isolated from natural sources, and has been synthesised by Jones *et al.* (1951, 1952); it is dehydrovitamin A_1.

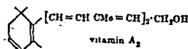


Jones *et al.* (1955) have also introduced a method for converting vitamin A_1 into vitamin A_2. Vitamin A_1 may be oxidised to vitamin A_1 aldehyde (**retinene$_1$**) by means of manganese dioxide in acetone solution (Morton *et al.*, 1948), and then treated as follows:



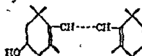




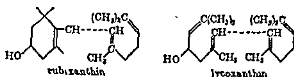


§8. Xanthophylls. The xanthophylls occur naturally, and all have the same carbon skeletons as the carotenes or lycopene (except flavoxanthin).

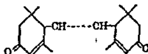
Cryptoxanthin, $C_{40}H_{56}O$, m.p. 169°, is monohydroxy-β-carotene; it has provitamin-A activity.



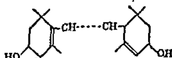
Rubixanthin, $C_{40}H_{56}O$, m.p. 160°, is monohydroxy-γ-carotene, and **lycoxanthin**, $C_{40}H_{56}O$, m.p. 168°, appears to be monohydroxylycopene



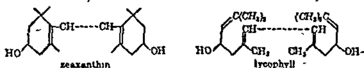
Rhodoxanthin, $C_{40}H_{52}O_2$, m.p. 219°, is believed to be the following diketone.



Lutein, $C_{40}H_{56}O_2$, m.p. 193°, was formerly known as xanthophyll; it is dihydroxy-α-carotene.



Zeaxanthin, m.p. 205°, and **lycophyll**, m.p. 179°, are the corresponding dihydroxy derivatives of β-carotene and lycopene, respectively.



§9. Carotenoid acids. These are compounds which do not contain 40 carbon atoms.

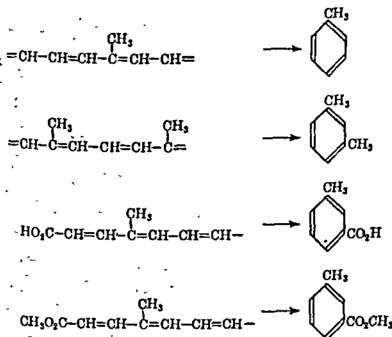
Bixin, $C_{25}H_{30}O_4$. Natural bixin is a brown solid, m.p. 198°, and is the *cis*-form; it is readily converted into the more stable *trans*-form, m.p. 216–217°.

When boiled with potassium hydroxide solution, bixin produces one molecule of methanol and a dipotassium salt which, on acidification, gives the dibasic acid **norbixin**, $C_{24}H_{28}O_4$. Thus bixin is a monomethyl ester, and can be esterified to give methylbixin.

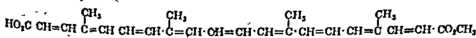
On catalytic hydrogenation, bixin is converted into perhydrobixin, $C_{25}H_{48}O_4$; thus there are 9 double bonds present in the molecule (Liebermann *et al.*, 1915). Perhydrobixin, on hydrolysis, forms perhydronorbixin. Oxidation of bixin with permanganate produces four molecules of acetic acid (Kuhn *et al.*, 1929); thus there are four —C(CH$_3$)= groups in the chain. Furthermore, since the parent hydrocarbon of perhydronorbixin, $C_{24}H_{46}O_4$, is $C_{22}H_{46}$

(the two carboxyl groups are regarded as substituents), the molecule is acyclic.

The thermal decomposition of bixin produces toluene, *m*-xylene, *m*-toluic acid and the methyl ester of this acid (Kuhn *et al.*, 1932). Hence the following assumptions may be made regarding the nature of the chain (*cf.* β-carotene, §3).



The foregoing facts may be explained by assuming the following structure for bixin (Kuhn *et al.*, 1932):

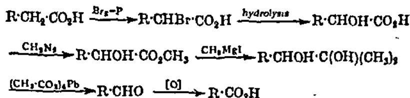


This structure is supported by the fact that perhydronorbixin has been synthesised, and shown to be identical with the compound obtained from the reduction of bixin (Karrer *et al.*, 1933).

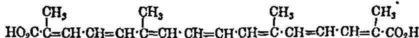
Crocetin, $C_{20}H_{24}O_4$. Crocetin occurs in saffron as the digentiobioside, *crocin*. The structure of crocetin was elucidated by Karrer *et al.* (1928) and Kuhn *et al.* (1931). Crocetin behaves as a dicarbox[illegible]
hydroger[illegible]
chromic [illegible]
of crocetin; thus there are 3–4 methyl side-chains. The structure

O

of crocetin was finally shown by the degradation of perhydronorbixin, $C_{24}H_{46}O_4$, by means of the following method:



This set of reactions was performed *twice* on perhydronorbixin, thereby resulting in the loss of four carbon atoms (two from each end); the product so obtained was perhydrocrocetin, $C_{20}H_{38}O_4$. On these results, crocetin is therefore:



This structure is supported by the fact that the removal of two carbon atoms from perhydrocrocetin by the above technique (one carbon atom is lost from each end) resulted in the formation of a *diketone*. The formation of this compound shows the presence of an α-methyl group at each end of the molecule. The structure of crocetin is further supported by the synthesis of perhydrocrocetin.

Reading References

Karrer and Jucker, *Carotenoids*, Elsevier (translated and revised by Braude, 1950).

Rodd (Ed.), *Chemistry of the Carbon Compounds*, Elsevier, Vol. IIA (1953). Ch. 10. The Carotenoid Group.

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley, Vol. IV (1953). Ch. 7. The Terpenes (see the section on Tetraterpenes).

Ann. Reports (Chem. Soc.), 1953, 50, 178. Carotenoids.

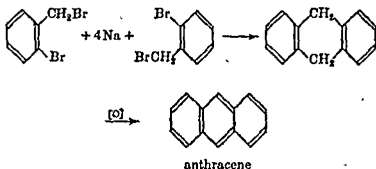
CHAPTER X

POLYCYCLIC AROMATIC HYDROCARBONS

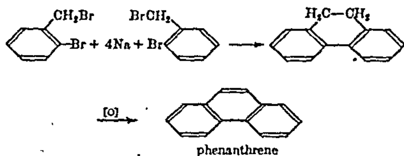
§1. Introduction. Naphthalene, anthracene, phenanthrene, fluorene, etc., have been described in Volume I. All these compounds occur in coal-tar, but also present are many polycyclic hydrocarbons containing four or more rings, and others of this type have been synthesised.

§2. General methods of preparation of polycyclic hydrocarbons. Before dealing with a number of individual hydrocarbons, it is instructive to review some of the general methods whereby these polycyclic hydrocarbons may be prepared (see also Vol. I).

(i) **Fittig reaction,** *e.g.*, anthracene and phenanthrene may be prepared by the action of sodium on *o*-bromobenzyl bromide.



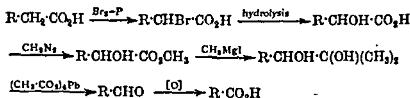
anthracene



phenanthrene

(ii) **Ullmann diaryl synthesis.** This method results in the

of crocetin was finally shown by the degradation of perhydronorbixin, $C_{24}H_{46}O_4$, by means of the following method:

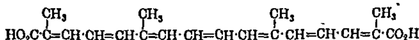


$$R{\cdot}CH_2{\cdot}CO_2H \xrightarrow{Br_2-P} R{\cdot}CHBr{\cdot}CO_2H \xrightarrow{hydrolysis} R{\cdot}CHOH{\cdot}CO_2H$$

$$\xrightarrow{CH_2N_2} R{\cdot}CHOH{\cdot}CO_2CH_3 \xrightarrow{CH_3MgI} R{\cdot}CHOH{\cdot}C(OH)(CH_3)_2$$

$$\xrightarrow{(CH_3{\cdot}CO_2)_4Pb} R{\cdot}CHO \xrightarrow{[O]} R{\cdot}CO_2H$$

This set of reactions was performed *twice* on perhydronorbixin, thereby resulting in the loss of four carbon atoms (two from each end); the product so obtained was perhydrocrocetin, $C_{20}H_{38}O_4$. On these results, crocetin is therefore:



$$HO_2C{\cdot}\overset{CH_3}{\overset{|}{C}}{=}CH{\cdot}CH{=}CH{\cdot}\overset{CH_3}{\overset{|}{C}}{=}CH{\cdot}CH{=}CH{\cdot}CH{=}\overset{CH_3}{\overset{|}{C}}{\cdot}CH{=}CH{\cdot}CH{=}\overset{CH_3}{\overset{|}{C}}{\cdot}CO_2H$$

This structure is supported by the fact that the removal of two carbon atoms from perhydrocrocetin by the above technique (one carbon atom is lost from each end) resulted in the formation of a *diketone*. The formation of this compound shows the presence of an α-methyl group at each end of the molecule. The structure of crocetin is further supported by the synthesis of perhydrocrocetin.

Reading References

Karrer and Jucker, *Carotenoids*, Elsevier (translated and revised by Braude, 1950).

Rodd (Ed.), *Chemistry of the Carbon Compounds*, Elsevier. Vol. IIA (1953). Ch. 10. The Carotenoid Group.

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol. IV (1953). Ch. 7. The Terpenes (see the section on Tetraterpenes).

Ann. Reports (Chem. Soc.), 1953, 50, 178. Carotenoids.

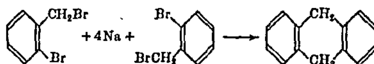
CHAPTER X

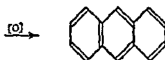
POLYCYCLIC AROMATIC HYDROCARBONS

§1. Introduction. Naphthalene, anthracene, phenanthrene, fluorene, etc., have been described in Volume I. All these compounds occur in coal-tar, but also present are many polycyclic hydrocarbons containing four or more rings, and others of this type have been synthesised.

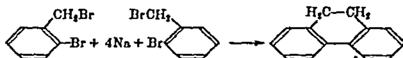
§2. General methods of preparation of polycyclic hydrocarbons. Before dealing with a number of individual hydrocarbons, it is instructive to review some of the general methods whereby these polycyclic hydrocarbons may be prepared (see also Vol. I).

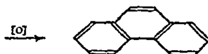
(i) **Fittig reaction,** *e.g.*, anthracene and phenanthrene may be prepared by the action of sodium on *o*-bromobenzyl bromide.





anthracene





phenanthrene

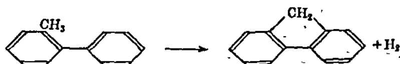
(ii) **Ullmann diaryl synthesis.** This method results in the

formation of isolated polynuclear compounds, *e.g.*, heating iodobenzene with copper powder in a sealed tube produces diphenyl

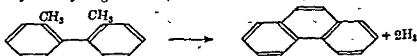


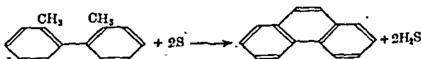
Compounds of the isolated system type can, under suitable conditions, be converted into condensed polycyclic compounds (see method iii). In certain cases, the Ullmann synthesis leads to condensed systems (see §4c).

(iii) Many compounds of the isolated system type can be converted into condensed systems by strong heating, *e.g.*, *o*-methyldiphenyl forms fluorene.

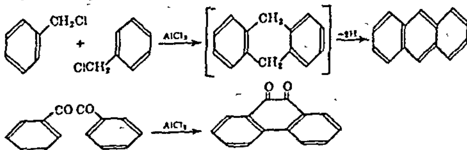


2:2'-Dimethyldiphenyl forms phenanthrene when passed through a red hot tube, but a much better yield is obtained when the dimethyldiphenyl is heated with sulphur. The latter is an example of cyclodehydrogenation (see also method vii).

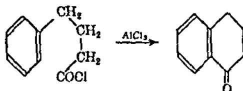




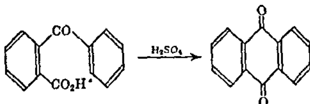
(iv) **Friedel-Crafts reaction.** Condensed polycyclic compounds may be prepared *via* an external or an internal Friedel-Crafts reaction. An example of the former is the preparation of anthracene from benzyl chloride; an example of the latter is the preparation of phenanthraquinone from benzil.



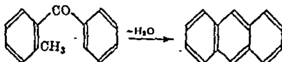
A very important case of the internal Friedel–Crafts reaction is that in which ring closure is effected on acid chlorides, *e.g.*, the conversion of γ-phenylbutyryl chloride to α-tetralone.



This type of ring closure may be effected by the action of concentrated sulphuric acid on the carboxylic acid itself, *e.g.*,

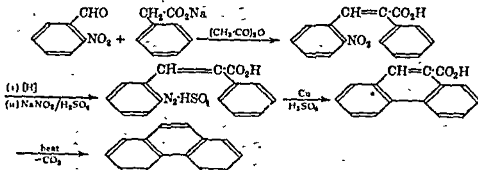


(v) **Elbs reaction.** In this method, polynuclear hydrocarbons are produced from a diaryl ketone containing a methyl group in the *o*-position to the keto group. The reaction is usually carried out by heating the ketone under reflux or at 400–450° until water is no longer evolved, *e.g.*, *o*-methylbenzophenone forms anthracene.

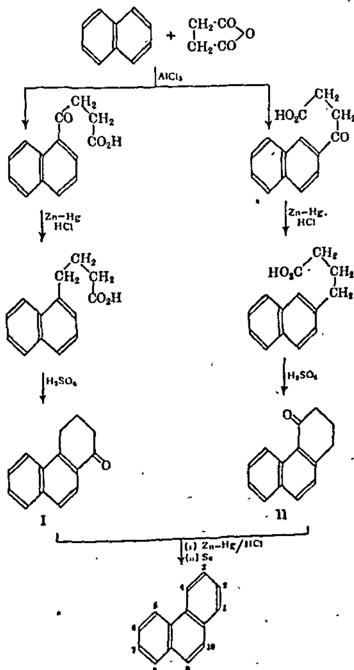


(vi) **Phenanthrene syntheses.** The phenanthrene nucleus is particularly important in steroid chemistry, and so a number of methods for synthesising phenanthrene are dealt with in some detail.

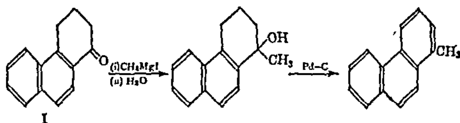
(*a*) **Pschorr synthesis** (1896). This method offers a means of preparing phenanthrene and substituted phenanthrenes with the substituents in known positions. Phenanthrene may be prepared as follows, starting with *o*-nitrobenzaldehyde and sodium β-phenylacetate.



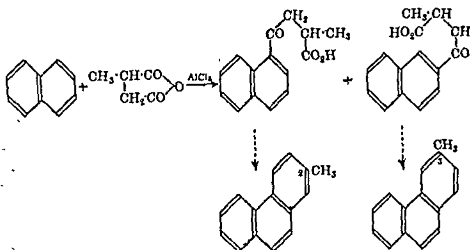
(*b*) **Haworth synthesis** (1932). Naphthalene is condensed with succinic anhydride in the presence of aluminium chloride in nitrobenzene solution. Two naphthoylpropionic acids are obtained, and these may be separated.



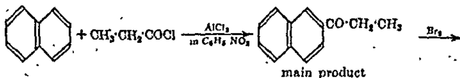
The Haworth synthesis is very useful for preparing alkylphenanthrenes with the alkyl group in position 1 (from I) or position 4 (from II); *e.g.*,

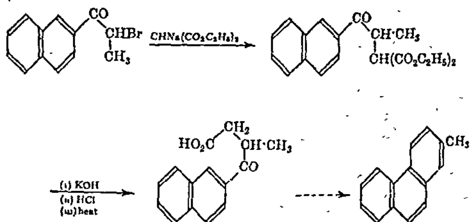


By using methylsuccinic anhydride instead of succinic anhydride, a methyl group can be introduced into the 2- or 3-position; in this case the condensation occurs at the less hindered keto group, *i.e.*, at the one which is farther removed from the methyl substituent.

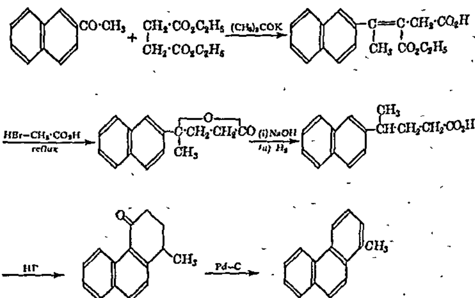


α-Bromoketone derivatives of naphthalene may be used in the malonic ester synthesis to prepare alkylphenanthrenes, *e.g.*,



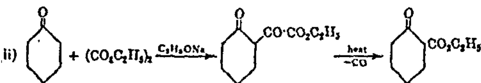


(*c*) **Stobbe condensation** (1893). This method has been improved by Johnson (1944), and has been used to prepare phenanthrene derivatives (see Vol. I); *e.g.*,

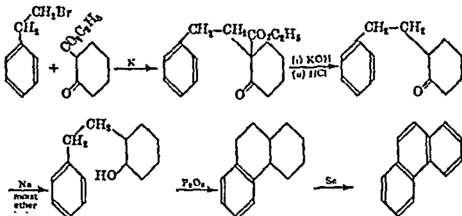


(*d*) **Bardhan–Sengupta synthesis** (1932). In this synthesis the starting materials are 2-phenylethyl bromide and ethyl *cyclo*hexane-2-carboxylate; these may be prepared as follows:

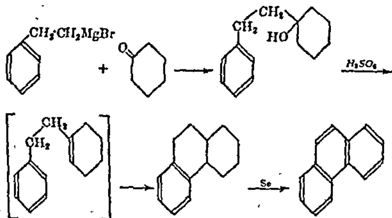




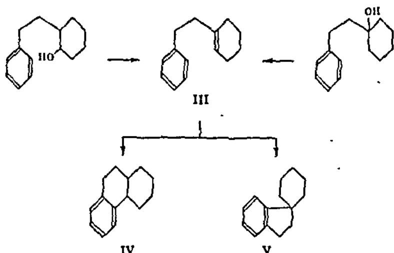
These two compounds are then treated as shown:



(*c*) **Bogert–Cook synthesis** (1933). The following chart shows the preparation of phenanthrene.



It might be noted here that the Bardhan-Sengupta and Bogert-Cook methods both proceed *via* the formation of olefin III, which then gives a mixture of octahydrophenanthrene IV and the spiran V.

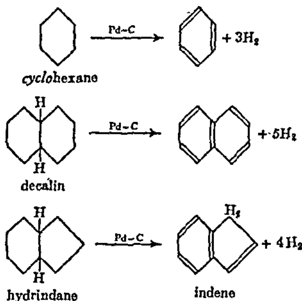


(vii) Dehydrogenation of hydroaromatic compounds with sulphur, selenium or palladised charcoal. This method is mainly confined to the dehydrogenation of six-membered rings, but five-membered rings may sometimes be dehydrogenated when they are fused to a six-membered ring. The general methods are as follows:

(*a*) Heating the compound with the calculated amount of sulphur at 200–220°; hydrogen is eliminated as hydrogen sulphide (Vesterberg, 1903).

(*b*) Heating the compound with the calculated amount of selenium at 250–280°; hydrogen is eliminated as hydrogen selenide (Diels, 1927).

(*c*) Heating the compound with palladium-charcoal up to about 300°, or passing the vapour of the compound over the catalyst heated at 180–350°; hydrogen is eliminated catalytically. Simple examples of catalytic dehydrogenation are:



Perhydro-compounds, *i.e.*, fully hydrogenated compounds, are readily dehydrogenated catalytically, but are very little affected, if at all, by the chemical reagents sulphur and selenium. Partially unsaturated compounds, however, are readily dehydrogenated by sulphur and selenium.

The method of dehydrogenation has been very useful in the elucidation of structure in terpene and steroid chemistry; specific examples are described in these two chapters. The following is an account of some of the general problems involved in dehydrogenation.

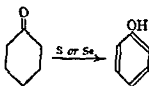
Originally, dehydrogenation was applied almost entirely to hydrocarbons, but subsequently it was found that many compounds containing certain functional groups could also be dehydrogenated, the nature of the products depending on the nature of the functional group.

(i) Alcoholic groups may be eliminated with the formation of unsaturated hydrocarbons, *e.g.*, eudesmol gives eudalene (§28b. VIII); cholesterol gives Diels' hydrocarbon (§1. XI).

(ii) Phenolic hydroxyl groups and methylated phenolic groups are usually unaffected by dehydrogenation with sulphur. With selenium, these groups may or may not be eliminated, but the higher the temperature at which the dehydrogenation is carried out (particularly above 300°), the greater is the likelihood of these groups being eliminated.

(iii) The products obtained from ketones depend on whether the

keto group is in a ring or in an open chain. Thus cyclic ketones are dehydrogenated to phenols, *e.g.*,

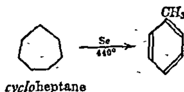


When the keto group is in a side-chain, then it is often unaffected.

(iv) Carboxyl (or carboalkoxyl) groups are eliminated when attached to a tertiary carbon atom, *e.g.*, abietic acid gives retene (§31. VIII). If, however, the carboxyl group is attached to a primary or secondary carbon atom, it is usually unaffected when the dehydrogenation is carried out with sulphur or palladium-charcoal. On the other hand, the carboxyl group is usually eliminated (decarboxylation) when selenium is used, but in some cases it is converted into a methyl group (see, *e.g.*, vitamin D, §6. XI).

(v) In a number of cases, dehydrogenation is accompanied by a rearrangement of the carbon skeleton, this tending to occur at higher temperatures and when the heating is prolonged.

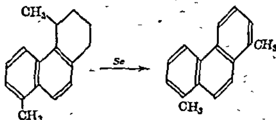
(*a*) Ring contraction may occur, *e.g.*,



*cyclo*heptane

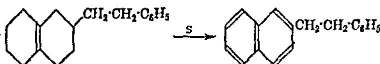
(*b*) Ring expansion may occur, *e.g.*, cholesterol gives chrysene (see §1. XI).

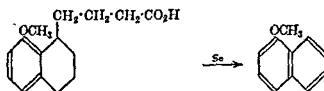
(*c*) Compounds containing an *angular* methyl group tend to eliminate this methyl group as CH_3SH or CH_3SeH, *e.g.*, eudesmol gives eudalene (§28b. VIII), cholesterol gives Diels' hydrocarbon (§1. XI). In some cases, the angular methyl group enters a ring,

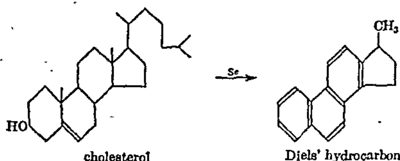


thereby bringing about ring expansion (*cf.* (*b*) above). On the other hand, a normal substituent methyl group may migrate to another position, *e.g.*, 5:6:7:8-tetrahydro-1.5-dimethylphenanthrene gives 1:8-dimethylphenanthrene on dehydrogenation with selenium.

(*d*) Side-chains larger than methyl may remain intact, or be eliminated, or be degraded, *e.g.*,



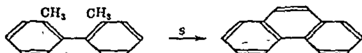




cholesterol Diels' hydrocarbon

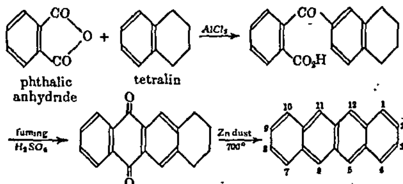
(*e*) Dehydrogenation may produce new rings (*cf.* method iii); *e.g.*



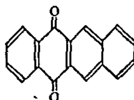


BENZANTHRACENES

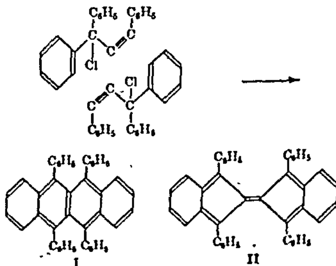
§3. Naphthacene (2.3-Benzanthracene), $C_{18}H_{12}$, is an orange solid, m.p. 357°. It occurs in coal-tar, and has been synthesised as follows (Fieser, 1931).



When oxidised with fuming nitric acid, naphthacene forms naphthacenequinone.

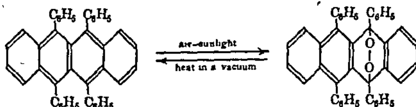


§3a. Rubrene (5:6:11:12-tetraphenylnaphthacene) may be prepared by heating 3-chloro-1:3:3-triphenylprop-1-yne alone, or better, with quinoline at 120° *in vacuo* (Dufraisse *et al.*, 1926).



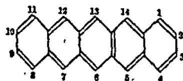
It is interesting to note that Dufraisse originally gave rubrene structure II, but changed it to I in 1935. The mechanism of the reaction is uncertain.

Rubrene is an orange-red solid, m.p. 334°. Its solution in benzene has a yellow fluorescence, but when this solution is shaken with air in sunlight, the fluorescence slowly disappears, and a white solid can now be isolated. This is rubrene peroxide, and when heated to 100–140° in a high vacuum, it emits yellow-green light and evolves oxygen, reforming rubrene.

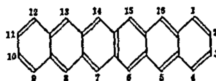


Rubrene peroxide is actually a derivative of 5:12-dihydronaphthacene, and so the molecule is not flat but folded about the O–O axis (the carbon atoms at 5 and 12 are tetrahedrally hybridised).

§3b. Two linear benzene derivatives of naphthacene have been prepared, *viz.*, **pentacene** (a deep violet-blue solid) and **hexacene** (a deep-green solid) (Clar, 1930, 1939).

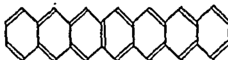


pentacene

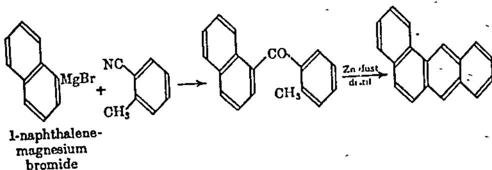


hexacene

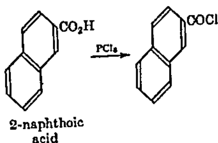
Clar (1942) thought he had prepared heptacene, but in 1950 he showed that the compound he had isolated was 1:2-benzohexacene. Various derivatives of heptacene, however, have been prepared.

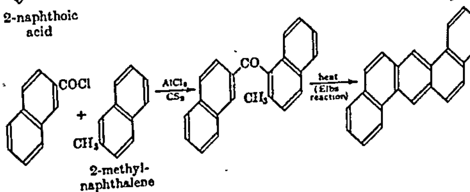

heptacene

§3c. **1:2-Benzanthracene**, m.p. 160°, occurs in coal-tar, and has been synthesised as follows (Bachmann, 1937).

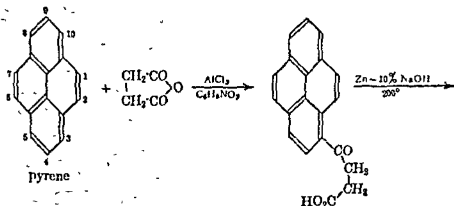


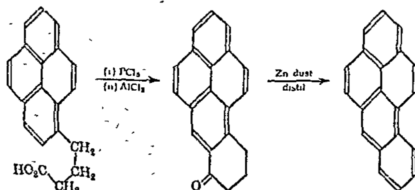
§3d. **1:2:5:6-Dibenzanthracene**, m.p. 266°, has been synthesised by Cook *et al.* (1931), who showed that it had strong carcinogenic activity.





§3e. **3:4-Benzpyrene** is a pale yellow solid, m.p. 179°, which is very strongly carcinogenic. It occurs in coal-tar, and has been synthesised as follows from pyrene (see §4b).





§3f. **20-Methylcholanthrene** is a pale yellow solid, m.p. 180°. It is a steroid derivative, and has been prepared by the degradation of, *e.g.*, cholesterol (see §3 iii. XI). Cook (1934) showed that methylcholanthrene has powerful carcinogenic properties, and Fieser *et al.* (1935) synthesised it in the following way :

$CH_3C_6H_4Cl + COCl{\cdot}CH_2{\cdot}CH_2Cl \xrightarrow[CS_2]{AlCl_3}$ $CH_3C_6H_3Cl{\cdot}CO{\cdot}CH_2{\cdot}CH_2Cl$ + $CH_3C_6H_3Cl{\cdot}CO{\cdot}CH_2{\cdot}CH_2Cl$

$\xrightarrow[105^\circ]{H_2SO_4}$ Cl, CH_3, CO, CH_2—CH_2 + Cl, CH_3, CH_2, CO—CH_2 $\xrightarrow[HCl]{Zn-Hg}$ Cl, CH_3

$\xrightarrow[245^\circ]{CuCN}$ CN, CH_3 $\xrightarrow[\text{bromide}]{\text{1-naphthylmagnesium}}$ CO, CH_3

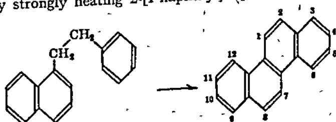
$\xrightarrow[\text{(Elbs reaction)}]{410^\circ}$ CH_3 or CH_3 (numbered 1–23)

The alternative way of writing the formula shows more clearly the relationship of methylcholanthrene to the steroids (see §3. XI for the method of numbering in cholesterol). The steroids are phenanthrene derivatives, and so methylcholanthrene may also be regarded as a phenanthrene derivative (instead of an anthracene derivative).

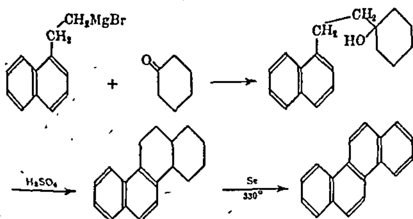
PHENANTHRENE DERIVATIVES

§4. **Chrysene** (1:2-benzphenanthrene) is a colourless solid, m.p. 251°. It occurs in coal-tar, and has been synthesised in several ways:

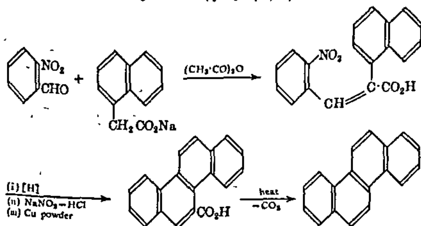
(i) By strongly heating 2-[1-naphthyl]-1-phenylethane.



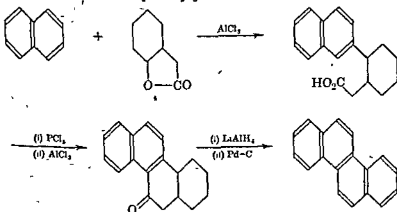
(ii) By a Bogert–Cook synthesis (*cf.* §2 (vi) *e*).



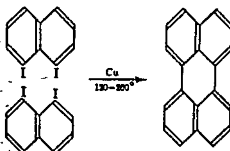
(iii) By a Pschorr synthesis (*cf.* §2 (vi) *a*).



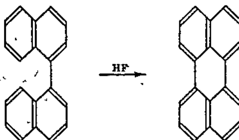
(iv) Phillips (1956) has prepared chrysene from naphthalene and the lactone of *trans* 2-hydroxy*cyclo*hexaneacetic acid:



(ii) Perylene may also be prepared by heating 1:8-di-iodonaphthalene with copper powder (*i.e.*, by an Ullmann synthesis; *cf.* §2. ii).



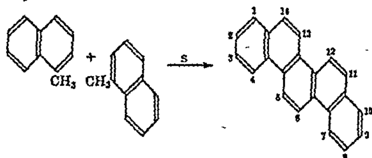
(iii) Perylene is formed when 1:1′-dinaphthyl is heated with hydrogen fluoride under pressure.



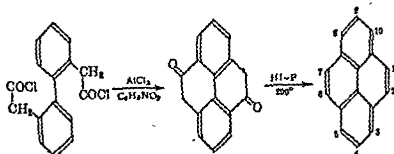
§4d. **Coronene,** m.p. 430°, is a yellow solid with a blue fluorescence in benzene solution; it has been found in coal-gas (Lindsay *et al.*, 1950). It was synthesised by Scholl *et al.* (1932), starting from *m*-xylene and anthraquinone-1:5-dicarbonyl chloride, the latter behaving in the tautomeric form shown in the following chart.

Chrysene is produced by the pyrolysis of indene, and also by the dehydrogenation of steroids with selenium.

§4a. **Picene** (1:2:7:8-dibenzphenanthrene), m.p. 365°, is obtained when cholesterol or cholic acid is dehydrogenated with selenium. It has been synthesised by heating 1-methylnaphthalene with sulphur at 300°.

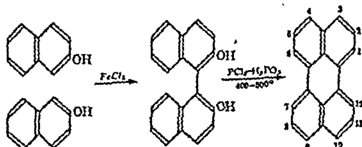


§4b. **Pyrene** is a colourless solid, m.p. 150°. It occurs in coal-tar, and has been synthesised from diphenyl-2:2'-diacetyl chloride as follows:

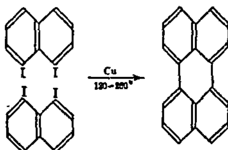


§4c. **Perylene** is a very pale yellow solid, m.p. 273°. It occurs in coal-tar, and has been synthesised in several ways.

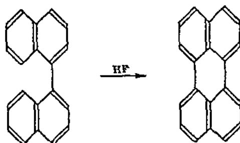
(i) 2-Naphthol, on treatment with ferric chloride solution, forms 1:1'-dinaphthol, and this, on heating with a mixture of phosphorus pentachloride and phosphorous acid, gives perylene.



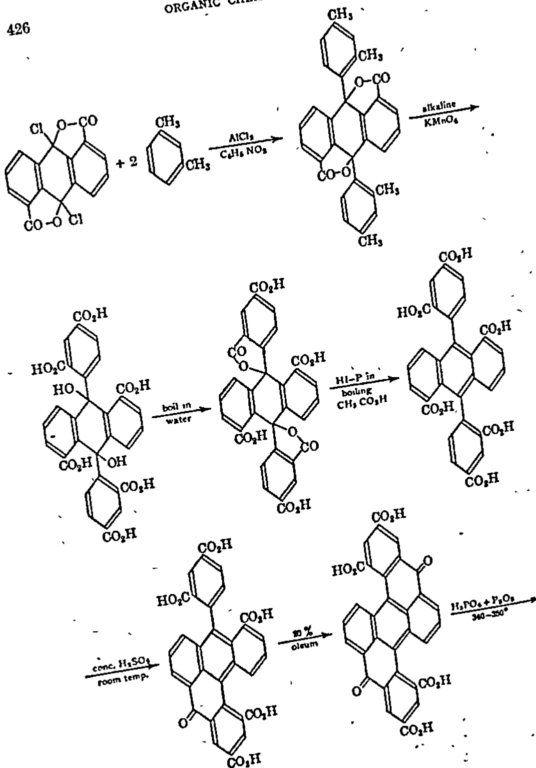
(ii) Perylene may also be prepared by heating 1:8-di-iodonaphthalene with copper powder (*i.e.*, by an Ullmann synthesis; *cf.* §2. ii).



(iii) Perylene is formed when 1:1′-dinaphthyl is heated with hydrogen fluoride under pressure.



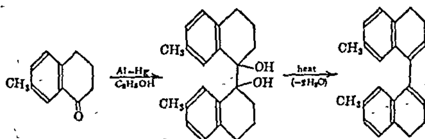
§4d. Coronene, m.p. 430°, is a yellow solid with a blue fluorescence in benzene solution; it has been found in coal-gas (Lindsay *et al.*, 1956). It was synthesised by Scholl *et al.* (1932), starting from *m*-xylene and anthraquinone-1:5-dicarbonyl chloride, the latter behaving in the tautomeric form shown in the following chart.

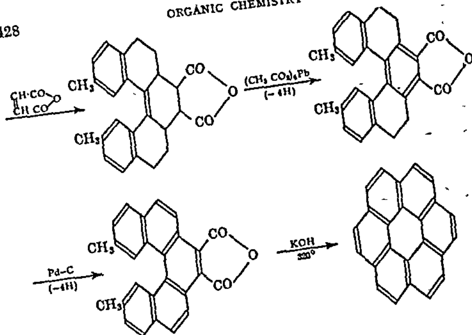

AlCl$_3$
C$_6$H$_5$NO$_2$
alkaline
KMnO$_4$
boil in
water
HI–P in
boiling
CH$_3$CO$_2$H
conc. H$_2$SO$_4$
room temp.
oleum
H$_3$PO$_4$ + P$_2$O$_5$
340–350°

CO_2H, O, O, O, O, CO_2H $\xrightarrow[180°]{HI-P}$ CO_2H, H_2, H_2, H_2, H_2, CO_2H $\xrightarrow[500°]{NaOH+Cu}$

$\xrightarrow[220°]{HNO_3}$ CO_2H CO_2H, CO_2H CO_2H $\xrightarrow[500°]{NaOH}$ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Newman (1940) has also synthesised coronene, starting from 7-methyltetralone, and proceeding as follows:





Reading References

Newer Methods of Preparative Organic Chemistry, Interscience Publishers (1948). Dehydrogenation with Sulphur, Selenium and Platinum Metals (pp. 21–59)

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol. IV (1953), pp. 1232– . Dehydrogenating Agents.

Genie, La Cyclodéshydrogénation Aromatique, *Ind. chim. belg.*, 1953, **18**, 670.

Ann. Reports (Chem. Soc.), 1936, **33**, 294. Dehydrogenation in the Determination of Structure.

Ann. Reports (Chem. Soc.), 1942, **39**, 165; 1953, **50**, 188. Polycyclic Aromatic Compounds.

Cook, Polycyclic Aromatic Hydrocarbons, *J.C.S.*, 1950, 1210.

Everest, *The Higher Coal-Tar Hydrocarbons*, Longmans, Green (1927).

Traité de Chimie Organique, Masson et Cie., Vol. XVII. Part II (1949).

Encyclopaedia of Organic Chemistry, Elsevier. Vol. 14 (1940). Tetracyclic and Higher-Cyclic Compounds. See also Vol. 14 Supplement (1951).

Cocker, Cross *et al.*, The Elimination of Non-angular Alkyl Groups in Aromatisation Reactions. Part II. *J.C.S.*, 1953, 2355.

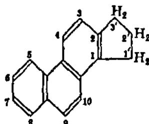
Cook (Ed.), *Progress in Organic Chemistry*, Butterworth. Vol. 2 (1953). Ch. 5. The Relationship of Natural Steroids to Carcinogenic Aromatic Compounds.

Badger, *The Structures and Reactions of Aromatic Compounds*, Cambridge Press (1954).

CHAPTER XI

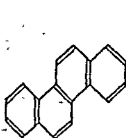
STEROIDS

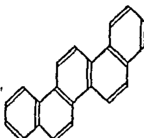
§1. Introduction. The steroids form a group of structurally related compounds which are widely distributed in animals and plants. Included in the steroids are the sterols (from which the name *steroid* is derived), vitamin D, the bile acids, a number of sex hormones, the adrenal cortex hormones, some carcinogenic hydrocarbons, certain sapogenins, etc. The structures of the steroids are based on the 1:2-*cyclo*pentenophenanthrene skeleton (Rosenheim and King, 1932; Wieland and Dane, 1932). All the



1:2-*cyclo*pentenophenanthrene

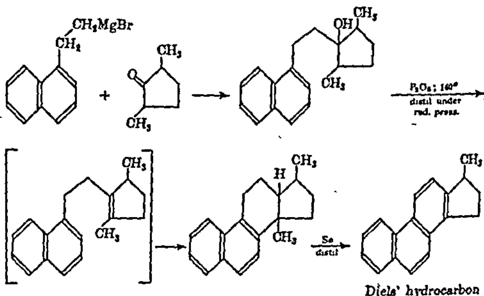
steroids give, among other products, Diels' hydrocarbon on dehydrogenation with selenium at 360° (Diels, 1927). In fact, a steroid could be defined as any compound which gives Diels' hydrocarbon when distilled with selenium. When the distillation with selenium is carried out at 420°, the steroids give mainly chrysene (§4. X) and a small amount of picene (§4a. X).


chrysene


picene

Diels' hydrocarbon is a solid, m.p. 126–127°. Its molecular formula is $C_{18}H_{16}$, and the results of oxidation experiments, X-ray

crystal analysis and absorption spectrum measurements showed that the hydrocarbon is probably 3′-methyl-1:2-*cyclo*penteno-phenanthrene. This structure for the compound was definitely established by synthesis, *e.g.*, that of Harper, Kon and Ruzicka (1934) who used the Bogert–Cook method (§2 (vi) *e*. X), starting from 2-(1-naphthyl)-ethylmagnesium bromide and 2:5-dimethyl*cyclo*pentanone.



Diels' *hydrocarbon*

STEROLS

§2. Sterols occur in animal and plant oils and fats. They are crystalline compounds, and contain an alcoholic group; they occur free or as esters of the higher fatty acids, and are isolated from the unsaponifiable portion of oils and fats. Cholesterol, cholestanol and coprostanol (coprosterol) are the animal sterols; ergosterol and stigmasterol are the principal plant sterols. The sterols that are obtained from animal sources are often referred to as the *zoosterols*, and those obtained from plant sources as the *phytosterols*. A third group of sterols, which are obtained from yeast and fungi, are referred to as the *mycosterols*. This classification, however, is not rigid, since some sterols are obtained from more than one of these groups.

§3. Cholesterol, $C_{27}H_{46}O$, m.p. 149°. This is the sterol of the higher animals, occurring free or as fatty esters in all animal cells, particularly in the brain and spinal cord. Cholesterol was first isolated from human gall-stones (these consist almost entirely of cholesterol). The main sources of cholesterol are the fish-liver oils, and the brain and spinal cord of cattle. Lanoline, the fat from wool, is a mixture of cholesteryl palmitate, stearate and oleate.

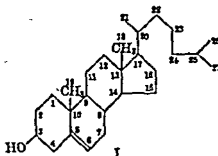
Cholesterol is a white crystalline solid which is optically active (lævorotatory). Cholesterol (and other sterols) gives many colour reactions, *e.g.*,

(i) *The Salkowski reaction* (1908). When concentrated sulphuric acid is added to a solution of cholesterol in chloroform, a red colour is produced in the chloroform layer.

(ii) *The Liebermann–Burchard reaction* (1885, 1890). A reddish-purple colour is developed when a solution of cholesterol in chloroform is treated with concentrated sulphuric acid and acetic anhydride.

When an ethanolic solution of cholesterol is treated with an ethanolic solution of digitonin (a saponin; see §19. iii), a large white precipitate of cholesterol digitonide is formed. This is a molecular complex containing one molecule of cholesterol and one of digitonin, from which the components may be recovered by dissolving the complex in pyridine (which brings about complete dissociation) and then adding ether (the cholesterol remains in solution and the digitonin is precipitated). Digitonide formation is used for the estimation of cholesterol.

The structure of cholesterol was elucidated only after a tremendous amount of work was done, particularly by Wieland, Windaus and their co-workers (1903–1932). Only a very bare outline is given here, and in order to appreciate the evidence that is going to be

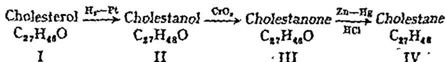


described, it is necessary to have the established structure of cholesterol at the beginning of our discussion. I is the structure of

cholesterol, and shows the method of numbering. The molecule consists of a *side-chain* and a *nucleus* which is composed of four rings; these rings are usually designated A, B, C and D or I, II, III and IV, beginning from the six-membered ring on the left (see also iii below). *It should be noted that the nucleus contains two angular methyl groups, one at* C_{10} *and the other at* C_{13}.

(i) **Structure of the ring system.** *Under this heading we shall deal with the nature of the ring system present in cholesterol; the problem of the angular methyl groups is dealt with later (see iv).*

The usual tests for functional groups showed that cholesterol contains *one double bond and one hydroxyl group. Now let us consider the following set of reactions.*



$$\underset{\text{I}}{\underset{C_{27}H_{46}O}{\text{Cholesterol}}} \xrightarrow{H_2-Pt} \underset{\text{II}}{\underset{C_{27}H_{48}O}{\textit{Cholestanol}}} \xrightarrow{CrO_3} \underset{\text{III}}{\underset{C_{27}H_{46}O}{\textit{Cholestanone}}} \xrightarrow[HCl]{Zn-Hg} \underset{\text{IV}}{\underset{C_{27}H_{48}}{\textit{Cholestane}}}$$

The conversion of cholesterol into cholestanol, II, shows the presence of one double bond in I, and the oxidation of II to the ketone cholestanone, III, shows that cholesterol is a secondary alcohol. Cholestane, IV, is a saturated hydrocarbon, and corresponds to the general formula C_nH_{2n-6}, and consequently *is tetracyclic; thus cholesterol is tetracyclic.*

When cholesterol is distilled with selenium at 360°, *Diels' hydrocarbon is obtained (see §1). The formation of this compound could be explained by assuming that this nucleus is present in cholesterol. The yield of this hydrocarbon, however, is always poor, and other products are always formed at the same time, particularly chrysene* (see §1). Thus, *on the basis of this dehydrogenation, the presence of the cyclopentenophenanthrene nucleus must be accepted with reserve.* Rosenheim and King (1932) thought that chrysene was the normal product of the selenium dehydrogenation, and so proposed (on this basis and also on some information obtained from X-ray analysis work of Bernal, 1932; see §4a) that the *steroids contained the chrysene skeleton. Within a few months, however,* Rosenheim and King (1932) modified this suggestion, as did also Wieland and Dane (1932). These two groups of workers proposed that the *cyclo*pentenophenanthrene nucleus is the one present *in cholesterol (i.e., in steroids in general).* This *structure fits far better all the evidence that has been obtained from a detailed investigation of the oxidation products of the sterols and bile acids.* This structure has now been confirmed by synthetic work, *e.g.*, the synthesis of equilenin (Bachmann *et al.*, 1940; see §17a), and by the synthesis of *cholesterol itself (see later in this section).*

Although an account of the oxidative degradation of the steroids cannot be discussed here, the following points in this connection are of some interest.

(i) The nature of the *nucleus* in sterols and bile acids was shown to be the same, since cholanic acid or *allo*cholanic acid is one of the oxidation products (see §4a for the significance of the prefix *allo*).

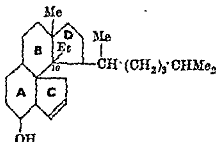
(ii) The oxidation of the bile acids led to the formation of products in which various rings were opened. The examination of these products showed that the positions of the hydroxyl groups were limited mainly to three positions, and further work showed that the hydroxyl groups behaved differently towards a given reagent, *e.g.*,

(*a*) The ease of oxidation of hydroxyl groups to keto groups by means of chromic acid is $C_7 > C_{12} > C_3$. More recently, Fonken *et al.* (1955) have shown that *tert.*-butyl hypochlorite apparently oxidises the 3-OH group selectively to the keto group; this reaction, however, failed with cholesterol. Sneedon *et al.* (1955) have also shown that the 3-OH group in steroids is oxidised by oxygen–platinum, but not those at 6, 7 or 12.

(*b*) The three keto groups are not equally readily reduced to a methylene group (by the Clemmensen reduction) or to an alcoholic group (by H_2—platinum). The ease of reduction is $C_3 > C_7 > C_{12}$. This is also the order for the ease of hydrolysis or acetylation when these positions are occupied by hydroxyl groups (see also testosterone, §13). More recently, it has been shown that the modified Wolff-Kishner reduction of Huang-Minlon (see Vol. I) on steroid ketones reduces keto groups at 3, 7, 12, 17 and 20, but not at 11. Another interesting point in this connection is that lithium aluminium hydride, in the presence of aluminium chloride, does not reduce unsaturated ketones to alcohols, *e.g.*, cholest-4-en-3-one, under these conditions, is reduced to cholest-3-ene (Broome *et al.*, 1956).

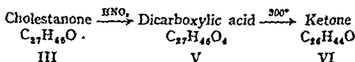
Thus a knowledge of (*a*) and (*b*) enabled workers to open the molecule at *different* points by oxidation under the appropriate conditions. This led to a large variety of degradation products, the examination of which enabled the nature of the nucleus to be ascertained.

(*c*) Blanc's rule was also used to determine the size of the various rings, but the failure of the rule in certain cases led to an erroneous formula; *e.g.*, ring C was originally believed to be five-membered. Thus Windaus and Wieland (1928) proposed the following formula for cholesterol, and the uncertain point (at that time) was the nature of the two extra carbon atoms. These were assumed to be present as an ethyl group at position 10, but Wieland *et al.* (1930) finally proved that there was no ethyl group at this posi-



These two " homeless " carbon atoms were not placed until Rosenheim and King first proposed that steroids contained the chrysene nucleus and then proposed the *cyclo*pentenophenanthrene nucleus (see above). Bernal (1932) also showed, from the X-ray analysis of cholesterol, ergosterol, etc., that the molecule was *thin*, whereas the above structure for the steroid nucleus would be rather thick.

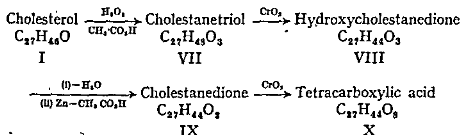
(ii) **Positions of the hydroxyl group and double bond.** Let us consider the following reactions:



Since the dicarboxylic acid V contains the same number of carbon atoms as the ketone (III) from which it is derived, the keto group in III must therefore be in a ring. Also, since pyrolysis of the dicarboxylic acid V produces a ketone with the loss of one carbon atom, it therefore follows from Blanc's rule that V is either a 1:6- or 1:7-dicarboxylic acid. Now we have seen that the nucleus contains three six-membered rings and one five-membered ring. Thus the dicarboxylic acid V must be obtained by the opening of ring A, B or C, and consequently it follows that the hydroxyl group in cholesterol (which was converted into the keto group in cholestanone; see (i) above) is in ring A, B or C.

Actually *two* isomeric dicarboxylic acids are obtained when cholestanone is oxidised. The formation of these two acids indicates that the keto group in cholestanone is flanked on either side by a methylene group, *i.e.*, the grouping $—CH_2·CO·CH_2—$ is present in cholestanone. Examination of the reference structure I of cholesterol shows that such an arrangement is possible only if the hydroxyl group is in ring A.

Now let us consider the further set of reactions:



In the conversion of I into VII, the double bond in I is hydroxylated. Since only two of the three hydroxyl groups in VII are oxidised to produce VIII, these two groups are secondary alcoholic groups (one of these being the secondary alcoholic group in cholesterol), and the third, being resistant to oxidation, is probably a tertiary alcoholic group. Dehydration of VIII (by heating *in vacuo*) and subsequent reduction of the double bond forms IX, and this, on oxidation, gives a tetracarboxylic acid *without loss of carbon atoms*. Thus the two keto groups in IX must be in *different* rings; had they been in the *same* ring, then carbon would have been lost and X not obtained. It therefore follows that the hydroxyl group and double bond in cholesterol must be in *different* rings. Furthermore, since IX forms a pyridazine derivative with hydrazine, IX is a γ-diketone. Since we have already tentatively placed the hydroxyl group in ring A, the above reactions can be readily explained if we place the hydroxyl group at position 3, and the double bond between 5 and 6. In the following equations only rings A and B are drawn; this is an accepted convention of focusing attention on any part of the steroid molecule that is under consideration (also note that full lines represent groups lying above the plane, and broken lines groups lying below the plane; see also §§4, 4a, 4b).

Noller (1939) has shown that the pyridazine derivative is a polymer, and so the interpretation that IX is a γ-diketone is rendered uncertain. Supporting evidence, however, for the above interpretation is afforded by the fact that when cholesterol is heated with copper oxide at 290°, cholestenone, XI, is produced, and this on oxidation with permanganate, forms a keto-acid, XII, with the loss of one carbon atom. The formation of XII indicates that the keto group and the double bond in cholestenone are in the *same* ring. The ultraviolet absorption spectrum of cholestenone shows that the keto group and the double bond are conjugated (Menschick *et al.*, 1932). These results can be explained if we assume that the double bond in cholesterol migrates in the formation of cholestenone, the simplest explanation being that the hydroxyl group is in position 3 and the

HO I → HO VII OH OH → O VIII OH O

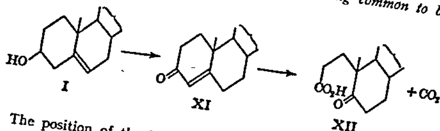
→ O IX H O → HO_2C HO_2C X H CO_2H CO_2H

IX ↓ N_2H_4

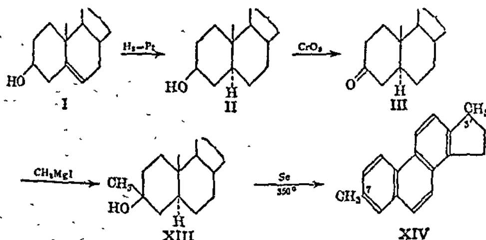
N H N

pyridazine derivative

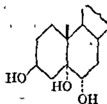
double bond between 5 and 6, *position 5 being common to both rings A and B.* Thus:



The position of the hydroxyl group at position 3 is definitely proved by the experiments of Kon *et al* (1937, 1939). These authors reduced cholesterol, I, to cholestanol, II, oxidised this to cholestanone, III, treated this with methylmagnesium iodide and dehydrogenated the product, a tertiary alcohol, XIII, to 3':7-dimethyl*cyclo*pentenophenanthrene, XIV, by means of selenium. The structure of XIV was proved by synthesis, and so the reactions may be formulated as follows, with the hydroxyl at position 3.



It might be noted here that the orientation of the two hydroxyl groups (introduced across the double bond in cholesterol) depends on the nature of the reagent used. With hydrogen peroxide, or *via* the oxide, the cholestanetriol is *trans*-5:6 (VII); with potassium permanganate or osmium tetroxide, the product is *cis*-5·6 (VII*a*; *cf.* §5a. IV).

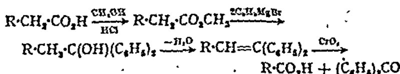


VII*a*

(iii) **Nature and position of the side-chain.** Acetylation of cholesterol produces cholesteryl acetate and this, on oxidation with chromium trioxide, forms a steam-volatile ketone and the acetate of a hydroxyketone (which is not steam volatile). The ketone was shown to be *iso*hexyl methyl ketone, $CH_3{\cdot}CO{\cdot}(CH_2)_3CH(CH_3)_2$. Thus this ketone is the side-chain of cholesterol, the point of attachment of the side-chain being at the carbon of the keto group. These results do not show where the side-chain is attached to the nucleus of cholesterol, but if we accept that the position is at 17, then we may formulate the reactions as follows:

The nature of the side-chain has also been shown by the application of the Barbier–Wieland degradation. Since this method also leads to evidence that shows *which ring* of the nucleus is attached to the side-chain, we shall consider the problem of the nature of the side-chain again.

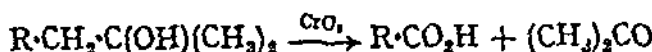
The Barbier–Wieland degradation offers a means of "stepping down" an acid one carbon atom at a time as follows:



$$R{\cdot}CH_2{\cdot}CO_2H \xrightarrow[HCl]{CH_3OH} R{\cdot}CH_2{\cdot}CO_2CH_3 \xrightarrow{2C_6H_5MgBr}$$

$$R{\cdot}CH_2{\cdot}C(OH)(C_6H_5)_2 \xrightarrow{-H_2O} R{\cdot}CH{=}C(C_6H_5)_2 \xrightarrow{CrO_3} R{\cdot}CO_2H + (C_6H_5)_2CO$$

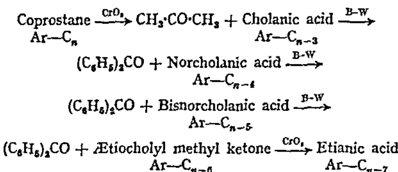
Methylmagnesium bromide may be used instead of phenylmagnesium bromide, and the alcohol so obtained may be directly oxidised:



$$R{\cdot}CH_2{\cdot}C(OH)(CH_3)_2 \xrightarrow{CrO_3} R{\cdot}CO_2H + (CH_3)_2CO$$

In the following account, only phenylmagnesium bromide will be used to demonstrate the application of the method to the steroids.

Cholesterol was first converted into coprostane (a stereoisomer of cholestane; see §§4, 4a). If we represent the nucleus of coprostane as Ar, and the side-chain as C_n, then we may formulate the degradation of coprostane as follows (B–W represents a Barbier-Wieland degradation):



$$\underset{Ar—C_n}{\text{Coprostane}} \xrightarrow{CrO_3} CH_3{\cdot}CO{\cdot}CH_3 + \underset{Ar—C_{n-3}}{\text{Cholanic acid}} \xrightarrow{B\text{-}W}$$

$$(C_6H_5)_2CO + \underset{Ar—C_{n-4}}{\text{Norcholanic acid}} \xrightarrow{B\text{-}W}$$

$$(C_6H_5)_2CO + \underset{Ar—C_{n-5}}{\text{Bisnorcholanic acid}} \xrightarrow{B\text{-}W}$$

$$(C_6H_5)_2CO + \underset{Ar—C_{n-6}}{\text{Ætiocholyl methyl ketone}} \xrightarrow{CrO_3} \underset{Ar—C_{n-7}}{\text{Etianic acid}}$$

The formation of acetone from coprostane indicates that the side-chain terminates in an *iso*propyl group. The conversion of bisnorcholanic acid into a ketone shows that there is an alkyl group on the α-carbon atom in the former compound. Furthermore, since the ketone is oxidised to etianic acid (formerly known as ætiocholanic acid) with the loss of one carbon atom, the ketone must be a methyl ketone, and so the alkyl group on the α-carbon atom in bisnorcholanic acid is a methyl group.

Now the carboxyl group in etianic acid is directly attached to the nucleus; this is shown by the following fact. When etianic acid is subjected to one more Barbier–Wieland degradation, a ketone, ætiocholanone, is obtained and this, on oxidation with nitric acid, gives a dicarboxylic acid, ætiobilianic acid, *without loss of any carbon atoms.* Thus ætiocholanone must be a *cyclic* ketone, and so it follows that there are *eight* carbon atoms in the side-chain, which must have the following structure in order to account for the foregoing degradations (see also the end of this section iii):

$$Ar \mid CH(CH_3) \mid CH_2 \mid CH_2 \mid CH_2 \mid CH(CH_3)_2$$

In addition to [illegible] there are also more recent methods [illegible]

(i) Gallagher *e*[illegible] method to eliminate *two* carbon atoms at a time:

$$Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}CO_2H \xrightarrow[(ii)\ CH_2N_2]{(i)\ SOCl_2} Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}CO{\cdot}CHN_2$$

$$\xrightarrow{HCl} Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}CO{\cdot}CH_2Cl \xrightarrow[AcOH]{Zn} Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}CO{\cdot}CH_3$$

$$\xrightarrow[(ii)\ -HBr]{(i)\ Br_2} Ar{\cdot}CHMe{\cdot}CH{=}CH{\cdot}CO{\cdot}CH_3 \xrightarrow{CrO_3} Ar{\cdot}CHMe{\cdot}CO_2H$$

(ii) Miescher *et al* (1944) have introduced a method to eliminate *three* carbon atoms at a time:

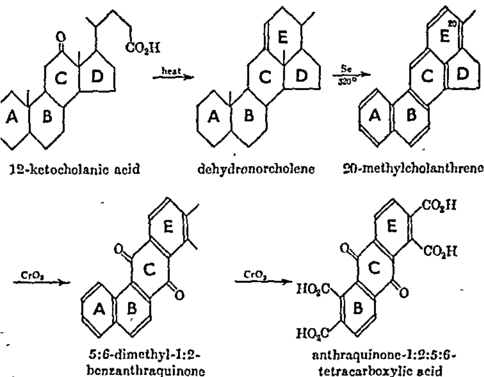
$$Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}CO_2Me \xrightarrow{2PhMgBr} Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH_2{\cdot}C(OH)Ph_2$$

$$\xrightarrow{-H_2O} Ar{\cdot}CHMe{\cdot}CH_2{\cdot}CH{=}CPh_2 \xrightarrow[succinimide]{N\text{-bromo-}} Ar{\cdot}CHMe{\cdot}CHBr{\cdot}CH{=}CPh_2$$

$$\xrightarrow{-HBr} Ar{\cdot}CMe{=}CH{\cdot}CH{=}CPh_2 \xrightarrow{CrO_3} Ar{\cdot}COMe$$

The problem now is: Where is the position of this side-chain? *This is partly answered by the following observation. The dicarboxylic acid, ætiobilianic acid, forms an anhydride when heated* with acetic anhydride. Thus the ketone (ætiocholanone) is probably a five-membered ring ketone (in accordance with Blanc's rule), and therefore the side-chain is attached to the five-membered ring *D*. The actual point of attachment to this ring, however, is not shown by this work. The formation of Diels' hydrocarbon (§1) from cholesterol suggests that the side-chain is at position 17, since selenium dehydrogenations may degrade a side-chain to a methyl group (see §2 vii. X). Position 17 is also supported by evidence obtained from X-ray photographs and surface film measurements. Finally, the following chemical evidence may be cited to show that the position of the side-chain is 17. As we have seen above, cholanic acid may be obtained by the oxidation of coprostane. Cholanic acid may also be obtained by the oxidation of deoxycholic acid (a bile acid; see §8) followed by a Clemmensen reduction. Thus the side-chains in cholesterol and deoxycholic acid are in the same position. Now deoxycholic acid can also be converted into 12-ketocholanic acid which, on heating to 320°, loses

water and carbon dioxide to form dehydronorcholene (Wieland *et al.*, 1930). This, when distilled with selenium, forms 20-methylcholanthrene, the structure of which is indicated by its oxidation to 5.6-dimethyl-1:2-benzanthraquinone which, in turn, gives on further oxidation, anthraquinone-1:2:5·6-tetracarboxylic acid (Cook, 1933). Finally, the structure of 20-methylcholanthrene has been confirmed by synthesis (Fieser *et al.*, 1935; see §3f. X). The foregoing facts can be explained only if the side-chain in cholesterol is in position 17; thus:



It should be noted that the isolation of methylcholanthrene affords additional evidence for the presence of the *cyclo*pentenophenanthrene nucleus in cholesterol.

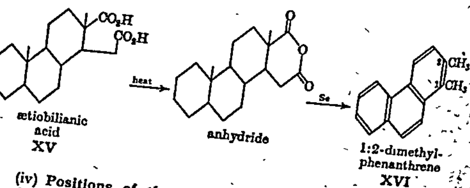
Thus, now that we know the nature and position of the side-chain, we can formulate the conversion of coprostane into ætiobilianic acid as follows:

coprostane $\xrightarrow{CrO_3}$ $CH_3 \cdot CO \cdot CH_3$ + cholanic acid $\xrightarrow{B-W}$

norcholanic acid $\xrightarrow{B-W}$ bisnorcholanic acid $\xrightarrow{B-W}$ ætiocholyl methyl ketone $\xrightarrow{CrO_3}$

etianic acid $\xrightarrow{B-W}$ ætiocholanone $\xrightarrow{HNO_3}$ ætiobilianic acid

A point of interest in this connection is that when the anhydride of ætiobilianic acid is distilled with selenium, 1:2-dimethylphenanthrene is obtained (Butenandt *et al.*, 1933). This also provides proof for the presence of the phenanthrene nucleus in cholesterol, and also evidence for the position of the C_{18} angular methyl group (see iv).

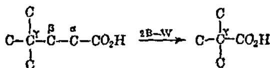


ætiobilianic acid XV — anhydride — 1:2-dimethylphenanthrene XVI

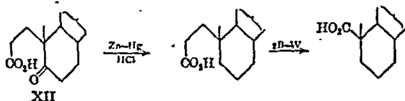
(iv) **Positions of the two angular methyl groups.** The *cyclo*pentenophenanthrene nucleus of cholesterol accounts for

seventeen carbon atoms, and the side-chain for eight. Thus twenty-five carbon atoms in all have been accounted for, but since the molecular formula of cholesterol is $C_{27}H_{46}O$, two more carbon atoms must be fitted into the structure. These two carbon atoms have been shown to be angular methyl groups.

In elucidating the positions of the hydroxyl group and double bond, one of the compounds obtained was the keto-acid XII. This compound, when subjected to the Clemmensen reduction and followed by two Barbier–Wieland degradations, gives an acid which is very difficult to esterify, and evolves carbon monoxide when warmed with concentrated sulphuric acid (Tschesche, 1932). Since these reactions are characteristic of an acid containing a carboxyl group attached to a tertiary carbon atom (*cf.* abietic acid, §31. VIII), the side-chain in XII must be of the type

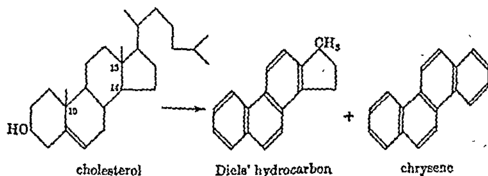


Thus there must be an alkyl group at position 10 in XII. This could be an ethyl group (as originally believed by Windaus and Wieland) or a methyl group, provided that in the latter case the second " missing " carbon atom can be accounted for. As we shall see later, there is also a methyl group at position 13, and so the alkyl group at position 10 must be a methyl group. On this basis, the degradation of XII may be formulated:



The position of the other angular methyl group is indicated by the following evidence. When cholesterol is distilled with selenium, chrysene is obtained as well as Diels' hydrocarbon (see §1). How, then, is the former produced if the latter is the ring skeleton of cholesterol? One possible explanation is that there is an angular methyl group at position 13, and on selenium dehydrogenation, this

methyl group enters the five-membered ring D to form a six-membered ring; thus:

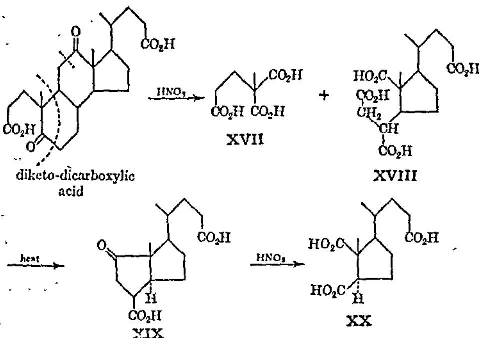


cholesterol → Diels' hydrocarbon + chrysene

This evidence, however, is not conclusive, since ring expansion could have taken place had the angular methyl group been at position 14. Further support for the positions of the two angular methyl groups is given by the following degradative experiments (Wieland *et al.*, 1924, 1928, 1933).

deoxycholic acid $\xrightarrow{CrO_3}$ dehydrodeoxycholic acid $\xrightarrow{HNO_3}$

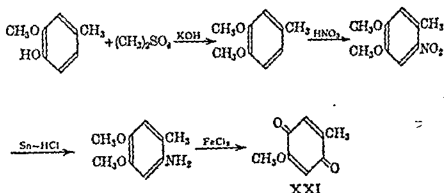
deoxybilianic acid $\xrightarrow[\text{in vacuo}]{\text{heat}}$ pyrodeoxybilianic acid $\xrightarrow{KMnO_4}$



XVII was shown to be butane-2:2.4-tricarboxylic acid; thus there is a methyl group at position 10. XVIII was shown to be a tetracarboxylic acid containing a *cyclo*pentane ring with a side-chain $—CH(CH_3)·CH_2·CH_2·CO_2H$. Thus this compound is derived from ring D. XX was also shown to be a tricarboxylic acid containing a *cyclo*pentane ring. Furthermore, one carboxyl group in XX was shown to be attached to a tertiary carbon atom, and so it follows that there is a methyl group at 13 or 14. XX was then shown to have the *trans* configuration, *i.e.*, the two carboxyl groups are *trans*. Thus its precursor XIX must have its two rings in the *trans* configuration (the methyl group and hydrogen atom at the junction of the rings are thus *trans*). Theoretical considerations of the strain involved in the *cis*- and *trans*-forms of XIX suggest that the *cis*-form of XIX would have been obtained had the methyl group been at position 14. Thus the position of this angular methyl group appears (from this evidence) to be at 13, and this is supported by the fact that ætiobilianic acid (XV, section iii) gives 1:2-dimethylphenanthrene (XVI) on dehydrogenation with selenium. Had the angular methyl group been at position 14, 1-methylphenanthrene would most likely have been obtained.

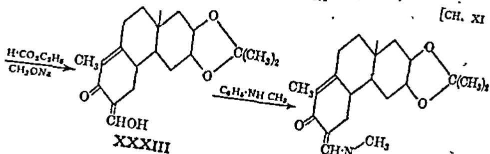
(v) **Synthesis of cholesterol.** Two groups of workers, *viz.*, Sir R. Robinson *et al.* (1951) and Woodward *et al.* (1951), have synthesised cholesterol. One of the outstanding difficulties

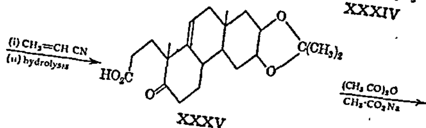
in the synthesis of steroids is the stereochemical problem. The cholesterol nucleus contains eight asymmetric carbon atoms and so 256 optical isomers are possible (see also §4 for further details). Thus every step in the synthesis which produced a new asymmetric carbon atom had to result in the formation of some (the more the better) of the desired stereoisomer, and at the same time resolution of racemic modifications also had to be practicable. Another difficulty was attacking a particular point in the molecule without affecting other parts. This problem led to the development of specific reagents. The following is an outline of the Woodward synthesis. 4-Methoxy-2·5-toluquinone, XXI, was prepared from 2-methoxy-*p*-cresol as follows:

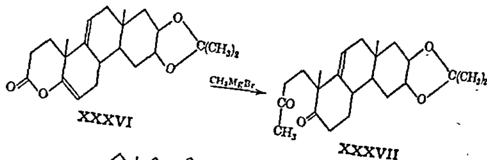


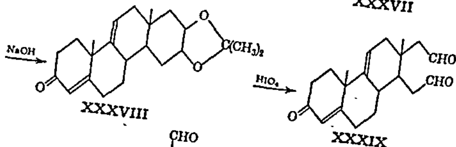
XXI was condensed with butadiene (Diels–Alder reaction) to give XXII. This had the *cis* configuration and was isomerised (quantitatively) to the *trans* isomer XXIII by dissolving in aqueous alkali, adding a seed crystal of the *trans* form, and then acidifying. XXIII, on reduction with lithium aluminium hydride, gave the glycol XXIV, and this, on treatment with aqueous acid, gave XXV. Conversion of XXV to XXVI by removal of the hydroxyl group was carried out by a new technique: XXV was acetylated and the product, the ketol acetate, was heated with zinc in acetic anhydride to give XXVI (reduction with metal and acid usually reduces $\alpha{\cdot}\beta$-unsaturated bonds in ketones). XXVI, on treatment with ethyl formate in the presence of sodium methoxide, gave the hydroxymethylene ketone XXVII (Claisen condensation). When this was treated with ethyl vinyl ketone in the presence of potassium *tert.*-butoxide, XXVIII was formed (Michael condensation). The object of the double bond in the ketone ring in XXVI is to prevent formylation occurring on that side of the keto group, and the purpose of the formyl group is to produce an active methylene group (this is now flanked on *both* sides by carbonyl groups). The

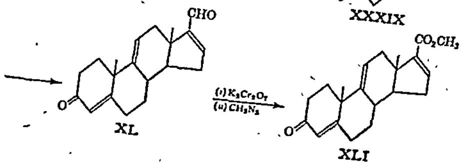
XLVI, on dehydration, gave one product, XLVII, and this, on catalytic hydrogenation (H_2—Pt), gave a mixture of cholestanyl acetates (the asymmetric C_{20} has been re-introduced). These acetates were separated and the desired isomer, on hydrolysis, gave cholestanol, XLVIII, which was identical with natural cholestanol. The conversion of cholestanol into cholesterol, LIII, is then carried out by a series of reactions introduced by various workers: XLVIII to XLIX (Bruce, 1943); XLIX to L (Butenandt *et al.*, 1935); L to LI (Ruzicka, 1938); LI to LII (Westphal, 1937); LII to LIII (Dauben *et al.*, 1950).


$H \cdot CO_2C_2H_5$
CH_3ONa
CH_3
O
O
$C(CH_3)_2$
O
CHOH
XXXIII
$C_6H_5 \cdot NH\ CH_3$
CH_3
O
O
$C(CH_3)_2$
O
$CH \cdot N$
CH_3
C_6H_5
XXXIV


(i) $CH_2=CH\ CN$
(ii) hydrolysis
HO_2C
O
O
$C(CH_3)_2$
O
XXXV
$(CH_3\ CO)_2O$
$CH_3 \cdot CO_2Na$


O
$C(CH_3)_2$
O
O
O
XXXVI
CH_3MgBr
O
$C(CH_3)_2$
O
CO
CH_3
O
XXXVII


NaOH
O
$C(CH_3)_2$
O
O
XXXVIII
HIO_4
CHO
CHO
O
XXXIX


CHO
O
XL
(i) $K_2Cr_2O_7$
(ii) CH_2N_2
CO_2CH_3
O
XLI

(i) [H]-form
(ii) H_2–Pt

CO_2CH_3

HO

XLII

CrO_3

CO_2CH_3

XLIII

(i) $NaBH_4$
(ii) hydrolysis

CO_2H

HO

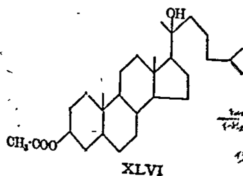
XLIV

(i) $(CH_3{\cdot}CO)_2O$
(ii) $SOCl_2$
(iii) $Cd(CH_3)_2$

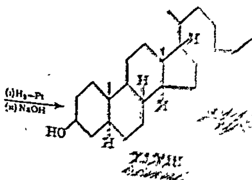
$CO{\cdot}CH_3$

$CH_3{\cdot}COO$

XLV



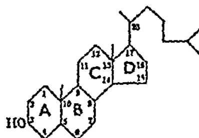
XLVI



$\xrightarrow{Br_2/HBr}$ L $\xrightarrow{\text{pyridine}}$ LI

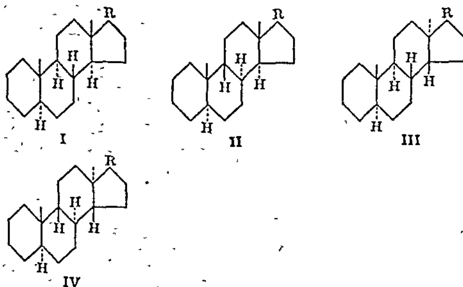
$\xrightarrow[(CH_3 \cdot CO)_2O]{CH_3 \cdot COCl}$ LII (CH_3COO) $\xrightarrow[\text{(ii) HCl}]{\text{(i) LiAlH}_4}$ LIII (HO) cholesterol

§4. **Stereochemistry of the steroids.** If we examine the fully saturated sterol, we find that there are eight dissimilar asymmetric carbon atoms in the nucleus (3, 5, 8, 9, 10, 13, 14 and 17). Thus there are $2^8 = 256$ optical isomers possible. If we also include the asymmetric carbon atom in the side-chain (20), then there are 512 optical isomers possible.



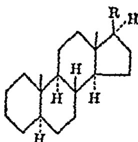
The stereoisomerism of the steroids is conveniently classified into two types, one dealing with the way in which the rings are fused together, and the other with the configurations of substituent groups, particularly those at C_3 and C_{17}.

§4a. Configuration of the nucleus. There are six asymmetric carbon atoms in the nucleus (5, 8, 9, 10, 13 and 14), and therefore there are $2^6 = 64$ optically active forms possible. X-ray analysis has shown that the steroid molecule is long and *thin*, *i.e.*, the molecule is essentially *flat* (Bernal, 1932). This is only possible if rings B and C are fused together in a *trans* manner (*cf. trans*-decalin, §11 vii. IV); rings A/B and C/D could be *cis* or *trans*. It has been found that all naturally occurring saturated steroids, except those of the heart poisons, belong either to the *cholestane series* or to the *coprostane series*; in the former the rings A/B are *trans*, and in the latter *cis*, the rings B/C and C/D being *trans* in *both* series. By convention a full line represents groups above the plane of the molecule, and a dotted (or broken) line represents groups below the plane (see also §11 vii. IV for conventions). Furthermore, *by convention*, the methyl group at C_{10} in *cholestane* has been placed *above* the plane of the molecule, and therefore this leads to *four* possible stereoisomers for cholestane (I–IV). X-ray analysis has



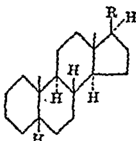
shown that the hydrogen atom at C_9 is *trans* to the methyl group at C_{10} (Bernal *et al.*, 1940), and this conclusion is supported by chemical evidence. Thus cholestane must be I or III. Further chemical work has shown that the methyl groups at C_{10} and C_{13} are *cis*, and so cholestane is I, and consequently coprostane is also I, except that in this case the hydrogen atom at C_5 is *above* the plane (rings A/B are *cis* in coprostane). The final point to be settled in connection with this problem of the configuration of cholestane is the

orientation of the side-chain R at C_{17}. Chemical evidence and X-ray analysis studies have shown that this side-chain is *above* the plane of the molecule (*i.e.*, *cis* with respect to the two angular methyl groups). Thus cholestane and coprostane are:



Cholestane

A/B *trans*
B/C *trans*
C/D *trans*
allo series



Coprostane

A/B *cis*
B/C *trans*
C/D *trans*
normal series

Compounds derived from cholestane are known as the ***allo*-compounds**, the prefix *allo* being reserved to indicate this configuration at C_5. Compounds derived from coprostane are known as the ***normal*-compounds**, but it should be noted that it is not customary to prefix compounds of this series by the word *normal*, e.g., *allo*cholanic acid can be derived from cholestane, whereas cholanic acid can be derived from coprostane.

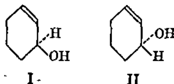
§4b. Configurations of substituent groups. The configuration of the side-chain at C_{17} has already been mentioned above. The only other configuration that we shall discuss here is that of the hydroxyl group at C_3. *By convention*, the hydroxyl at C_3 in cholestanol (and cholesterol) is taken as being above the plane of the ring, *i.e.*, the hydroxyl group is taken as being in the *cis* position with respect to the methyl group at C_{10}. This configuration occurs in all *natural* sterols, and gives rise to the **β-series**, the prefix *β* always indicating that the substituent group lies *above* the plane of the molecule. When the hydroxyl group lies below the plane, the compounds are said to belong to the **α- or epi series**; the prefix epi indicates the *epimer* due to the inversion of the configuration of C_3.

X-ray analysis studies have shown that the hydroxyl group in cholesterol is above the plane of the molecule, *i.e.*, it is *cis* to the

methyl group at C_{10}. This has been confirmed by chemical evidence (Shoppee, 1947).

The assignment of the configurations of C_7 and C_{17} in steroid alcohols has been determined by Prelog *et al.* (1953) by arguments based on asymmetric syntheses (see §7. III). It has been shown that the configuration of the hydroxyl group in, *e.g.*, cholestan-7α-ol and androsten-17β-ol is in agreement with the accepted conventional steroid formula.

Mills (1952) has also correlated the configurations of steroids with glyceraldehyde. This author collected the molecular optical rotations of a number of pairs of epimeric *cyclo*hex-2-enols and their esters, and on the assumption that the configurations given (in the literature) were correct, Miles showed that the alcohol represented as I is more lævorotatory than its epimer II. The differences



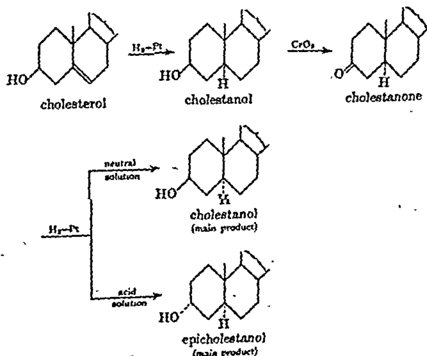
in rotation are large, and are increased on esterification. Mills then applied this rule to seven known pairs of epimeric, allylic steroid alcohols, and found that the differences were those which may be predicted on the basis that the conventional steroid formulæ represent the absolute configurations. Thus the configuration of the 3β-hydroxyl group in cholesterol corresponds to that of D(+)-glyceraldehyde.

These stereochemical relationships of steroids to D(+)-glyceraldehyde have now been proved by the degradation of cholesterol to derivatives of (+)-citronellal (§23e. VIII), in which the only asymmetric carbon atom is the C_{20} of the steroid (Cornforth *et al.*, 1954; Riniker *et al.*, 1954). Thus the arbitrary choice of placing the angular methyl groups above the plane in the cholesterol nucleus (*i.e.*, the β-configuration) has proved to be the absolute configuration. Furthermore, since the configuration of the 3-hydroxyl group in cholesterol is β, this configuration is also the absolute one.

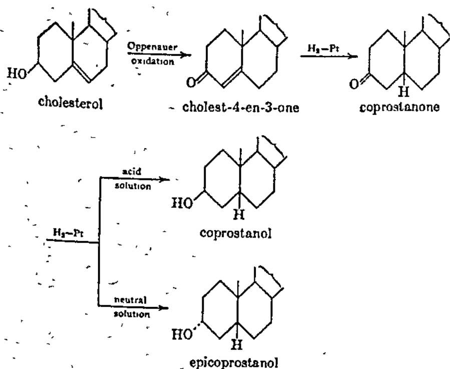
The use of optical rotations to determine structure and configuration in steroid chemistry was begun by Callow *et al.* (1936) who observed that all natural steroids with a 5:6-double bond are lævorotatory. These authors also observed that in sterols, bile acids and androstane compounds, inversion at C_3 from the β- to the α-configuration usually produces a slightly increased dextrorotation. As pointed out above, this approach has been used by Mills.

Barton (1944–) has also applied the method of optical rotations to steroid chemistry, and has called his treatment the Method of Molecular Rotation Differences (*this is a modification of the Rule of Shift*, §12 I). The basis of this method is that the molecular rotation of any steroid is considered as the sum of the rotation of the fundamental structure (which is the parent hydrocarbon cholestane, androstane, or pregnane) and the rotations contributed by the functional groups (these are called the Δ values). The Δ value of a given group is a characteristic of its position and orientation, and the Δ values of different groups are independent of one another provided that unsaturated groups are not present, *i.e.*, conjugation is absent, or that the groups are not too close together, *i.e.*, are separated by 3 or 4 saturated carbon atoms. In this way it has been possible to assign configurations and also the positions of double bonds.

§4c. The preparation of the "stanols". The catalytic hydrogenation (platinum) of cholesterol (cholest-5-en-3β-ol) produces only cholestanol (cholestan-3β-ol). On the other hand, oxidation of cholestanol with chromium trioxide in acetic acid gives cholestanone and this, on catalytic reduction in *neutral* solution, gives mainly cholestanol, whereas catalytic reduction in *acid* solution gives mainly epicholestanol (cholestan-3α-ol).



The corresponding C_5 epimers, coprostanol (coprostan-3β-ol) and epicoprostanol (coprostan-3α-ol), may be prepared from cholesterol as follows, the first step being the conversion of cholesterol into cholest-4-en-3-one by means of the Oppenauer oxidation (aluminium *tert.*-butoxide in acetone; see also Vol. I).



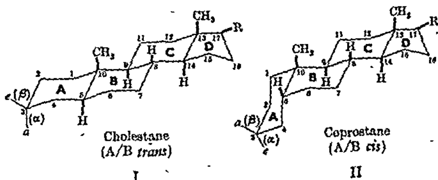
A detailed study of the catalytic reduction of the decalones has shown that in an acid medium the product is usually the *cis*-compound, whereas in a neutral or alkaline medium the product is usually the *trans*-compound (von Auwers, 1920; Skita, 1920). This principle, which is known as the *Auwers–Skita rule of catalytic hydrogenation*, was used by Ruzicka (1934) to determine the configurations of the above "stanols". The configurations assigned have been supported by measurement of the rates of hydrolysis of the acetates of the various "stanols" (Ruzicka *et al.*, 1938). The acetates of cholestanol and epicoprostanol are hydrolysed much faster than those of epicholestanol and coprostanol (see §4d).

Now let us consider the configuration at C_5. The results of experiments on the catalytic hydrogenation of substituted *cyclo*hexanones and substituted phenols have led to the generalisation that the initial addition is *cis*, and occurs on the more accessible side of the double bond (Peppiatt *et al.*, 1955; Wicker, 1956). In

accordance with this generalisation, it has been found that when saturated steroids of the A/B-*cis*- and the A/B-*trans*- series are produced by catalytic hydrogenation of 3α-substituted Δ^5-steroids, then the larger the size of the 3α-substituent, the larger is the proportion of the A/B-*cis*-steroid; in some cases, this *cis*-steroid is apparently formed exclusively (Shoppee *et al.*, 1955).

§4d. Conformational analysis of steroids. The Auwers-Skita rule of catalytic hydrogenation (§4c) cannot be used with certainty since, as pointed out, the product is *usually* mainly *cis* or *trans* according to the conditions, and hence the exceptions can only be ascertained as such by other evidence. Barton (1953) has restated this Auwers-Skita rule of catalytic hydrogenation as follows: *Catalytic hydrogenation of ketones in strongly acid media (rapid hydrogenation) produces the axial hydroxyl compound, whereas hydrogenation in neutral media (slow hydrogenation) produces the equatorial alcohol if the ketone is unhindered or the axial alcohol if the ketone is very much hindered.*

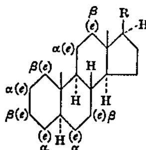
All the evidence obtained has shown that all the *cyclo*hexane rings in the steroid nucleus are chair forms; thus I is cholestane, and II is coprostane.



The effect of conformation on the course and rate of reactions has been discussed in §12. IV. The following is a summary of the generalisations that have been formulated:

(i) Equatorial groups are normally more stable than axial. Thus, when a (polycyclic) secondary alcohol is equilibrated with alkali, it is the equatorial isomer that predominates in the product. Similarly, when a (polycyclic) ketone is reduced with sodium and ethanol, the predominant isomer in the product is the equatorial alcohol (the more stable form). Furthermore, because of the rigidity of the system (which prevents interconversion of chair forms), the stable configurations of hydroxyl groups at different

positions in the cholestane series will be as shown in III (compare this with I).



III

The following are examples of equilibration (using sodium ethoxide at 180°):

Cholestanol [3β(e)] ⇌ Epicholestanol [3α(a)] (10% ↓, 90% ↑)

Coprostanol [3β(a)] ⇌ Epicoprostanol [3α(e)] (10% ↑, 90% ↓)

(ii) Equatorial hydroxyl and carboxyl groups are esterified more rapidly than the corresponding axial groups. Similarly, hydrolysis of equatorial esters and acyloxy groups is more rapid than for the corresponding axial isomers. These principles explain Ruzicka's results on the " stanols " (§4c); in the acetates of cholestanol and epicoprostanol, the acetoxy groups are equatorial, whereas in the acetates of epicholestanol and coprostanol these groups are axial and therefore subject to 1:3-interactions. Hence the former pair are hydrolysed more rapidly than the latter pair.

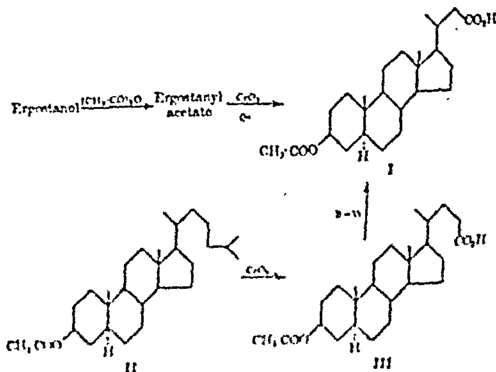
Empirical methods, using infra-red spectra, have been developed by Jones *et al.* (1951, 1952) for determining the conformation of 3-hydroxy (and 3-acetoxy) steroids; characteristic bands are given by the axial and equatorial groups.

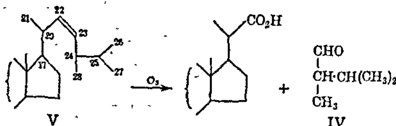
(iii) Secondary axial alcohols are more rapidly oxidised by chromic acid (or hypobromous acid) than secondary equatorial alcohols. Schreiber *et al.* (1955) have shown that the more hindered the alcohol, the faster is the oxidation (with chromic acid).

(iv) Bimolecular ionic elimination reactions occur readily when the two groups (which are eliminated) are *trans*-diaxial, and less readily when *trans*-diequatorial or *cis*-axial : equatorial.

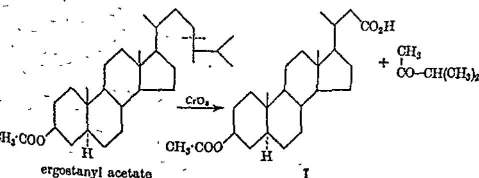
(v) Epoxides are attacked by, *e.g.*, hydrogen bromide, to give the *trans*-diaxial compound. Reduction with lithium aluminium hydride or catalytic hydrogenation converts epoxides into the axial hydroxy compound.

§5. **Ergosterol**, $C_{28}H_{44}O$, m.p. 163°, occurs in yeast. Ergosterol forms esters, *e.g.*, an acetate with acetic anhydride; thus there is a hydroxyl group present in ergosterol. Catalytic hydrogenation (platinum) of ergosterol produces ergostanol, $C_{28}H_{50}O$; thus there are three double bonds in ergosterol. When ergostanol is acetylated and the product then oxidised, the acetate of 3β-hydroxy-nor*allo*cholanic acid, I, is obtained (Fernholz *et al.*, 1934). The identity of I is established by the fact that cholestanyl acetate, II (a compound of known structure), gives, on oxidation, the acetate of 3β-hydroxy*allo*cholanic acid, III, and this, after one Barbier-Wieland degradation (§3 iii), gives I; thus:



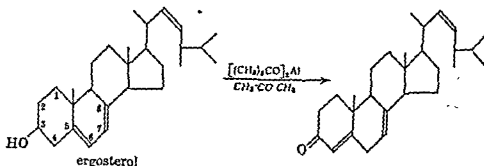


On this basis, the oxidation of ergostanyl acetate to the acetate of 3β-hydroxynor*allo*cholanic acid, I, is readily explained.



We have now accounted for all the structural features of ergosterol except the positions of the three double bonds. The position of one of these is actually shown in the above account; it is C_{22}—C_{23}. The side-chain must contain only *one* double bond, since if more than one were present, more than one fragment (IV) would have been removed on ozonolysis. Thus the other two double bonds must be in the nucleus. When heated with maleic anhydride at 135°, ergosterol forms an adduct, and so it follows that the two double bonds (in the nucleus) are conjugated (Windaus *et al.*, 1931). Now ergosterol has an absorption maximum at 2,810 A. Conjugated *acyclic* dienes absorb in the region of 2,200–2,500 A, but if the diene is in a *ring system*, then the absorption is shifted to the region 2,600–2,900 A. Thus the two double bonds in the nucleus of ergosterol are in *one* of the rings (Dimroth *et al.*, 1936). When ergosterol is subjected to the Oppenauer oxidation (aluminium *tert*.-butoxide and acetone), the product is an $\alpha{:}\beta$-unsaturated ketone (as shown from its absorption spectrum). This can only be explained by assuming that one of the double bonds is in the 5:6-position, and moves to the 4:5-position during the oxidation (*cf.* cholesterol, §3 ii). The other double bond is therefore 7:8 in order to be

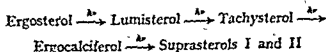
conjugated with the one that is 5:6. Thus the conjugated system is in ring B and the oxidation is explained as follows:



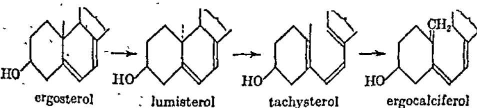
§6. **Vitamin D.** This vitamin is the antirachitic vitamin; it is essential for bone formation, its function being the control of calcium and phosphorus metabolism.

Steenbock *et al.* (1924) showed that when various foods were irradiated with ultraviolet light, they acquired antirachitic properties. This was then followed by the discovery that the active compound was in the unsaponifiable fraction (the sterol fraction). At first, it was believed that the precursor of the active compound was cholesterol, but subsequently the precursor was shown to be some "impurity" that was in the cholesterol fraction (*e.g.*, by Heilbron *et al.*, 1926). The ultraviolet absorption spectrum of this "impure cholesterol" indicated the presence of a small amount of some substance that was more unsaturated than cholesterol. This led to the suggestion that ergosterol was the provitamin D in the "impure cholesterol", and the investigation of the effect of ultraviolet light on ergosterol resulted in the isolation from the irradiated product of a compound which had very strong antirachitic properties. This compound was named **calciferol** by the Medical Research Council (1931), and **vitamin D_1** by Windaus (1931). This potent crystalline compound, however, was subsequently shown to be a molecular compound of calciferol and lumisterol (one molecule of each). Windaus (1932) therefore renamed the pure potent compound as **vitamin D_2**, but the M.R.C. retained the original name calciferol. The *Chemical Society* (1951) has proposed the name **ergocalciferol** for this pure compound.

When irradiated with ultraviolet light, ergosterol undergoes a number of changes, which are as follows:



Only ergocalciferol has antirachitic activity. The changes that take place during the irradiation appear to affect only ring B and the angular methyl group at C_{10} (the structure of tachysterol is uncertain, and the structures of the suprasterols more uncertain). The following chart shows the changes that are believed to occur:



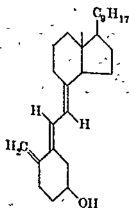
§6a. **Ergocalciferol** (calciferol, vitamin D_2) is an optically active crystalline solid, m.p. 115–117°. Its molecular formula is $C_{28}H_{44}O$, and since it forms esters, the oxygen is present as a hydroxyl group. Furthermore, since ergocalciferol gives a ketone on oxidation, this hydroxyl group is a secondary alcoholic group. Ozonolysis of ergocalciferol produces, among other products, methyl*iso*propylacetaldehyde. Thus the side-chain in ergocalciferol is the same as that in ergosterol. Catalytic hydrogenation converts ergocalciferol into the fully saturated compound octahydroergocalciferol, $C_{28}H_{52}O$. This shows that there are four double bonds present, and since one is in the side-chain, three are therefore in the nucleus. The parent hydrocarbon of ergocalciferol is $C_{28}H_{52}$, and since this corresponds to the general formula C_nH_{2n-4}, the molecule therefore is *tricyclic*. Furthermore, ergocalciferol does not give Diels' hydrocarbon when distilled with selenium. These facts indicate that ergocalciferol does not contain the four-ring system of ergosterol. The problem is thus to ascertain which of the rings in ergosterol has been opened in the formation of ergocalciferol. The following reactions of ergocalciferol are readily explained on the assumption that its structure is I. The absorption spectrum of the semicarbazone of II ($C_{21}H_{34}O$) was shown to be characteristic of α:β-unsaturated aldehydes. The absence of the hydroxyl group and the carbon content of II indicate the *absence* of ring A. These facts suggest that in ergocalciferol "ring B" is open between C_9 and C_{10}, and that II arises by scission of the molecule at a double bond in position 5:6, and can be an α:β-unsaturated aldehyde only if there is a double bond at 7:8 (these double bonds are also present in ergosterol). The isolation of the ketone III ($C_{19}H_{32}O$) confirms the presence of the double bond at 7.8 (Heilbron *et al.*, 1935).

The isolation of formaldehyde (IV) shows the presence of an

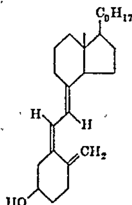
of VII, the structure of which is shown by the products VIII, IX, X and XI (Windaus *et al.*, 1936). The production of 2:3-dimethylnaphthalene (VIII) is in keeping with the fact that carboxyl groups sometimes give rise to methyl groups on selenium dehydrogenation (*cf.* §2 vii. X). Similarly, the formation of naphthalene, IX, and naphthalene-2-carboxylic acid, X, shows the presence of rings A and "B" in VII. Catalytic reduction of VII (to reduce the double bond in the *side-chain* only), followed by ozonolysis, gives XI. Thus the formation of these compounds VIII–XI establishes the structure of VII, and shows that the double bonds are at 5 6, 10 19, and 7:8.

X-ray analysis studies of the 4-iodo-3-nitrobenzoate of ergocalciferol confirm structure I for ergocalciferol (Crowfoot *et al.*, 1948).

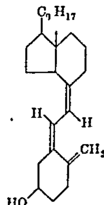
The presence of the two double bonds 5:6 and 7.8 gives rise to the possibility of various geometrical isomeric forms for ergocalciferol. When ergosterol is irradiated with ultraviolet light, it appears that pre-ergocalciferol is the final product, and this is then readily converted into the vitamin by gentle warming (Velluz *et al.*, 1948, 1949), the two being in equilibrium, with the vitamin predominating. Sondheimer *et al.* (1955), from a study of the ultraviolet spectra, have suggested that ergocalciferol has configuration XII, and pre-ergocalciferol XIII. On the other hand, Braude *et al.* (1955) have proposed XIII for ergocalciferol, and



XII



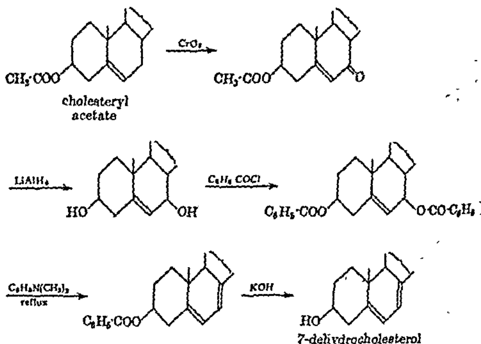
XIII



XIV

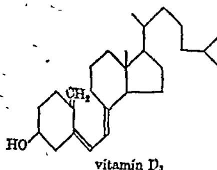
XIV for pre-ergocalciferol. Crowfoot *et al.* (1957), from calculations of electron densities in the ester crystal (the 4-iodo-3-nitrobenzoate), have shown that their results agree with XIII being the stereochemical arrangement in ergocalciferol.

§6b. Vitamins D_3 and D_4. A detailed biological investigation has shown that the vitamin D in cod-liver oil is not identical with ergocalciferol, and that vitamin D activity could be conferred on cholesterol, or on some impurity in cholesterol other than ergosterol. Windaus (1935) therefore suggested that natural vitamin D (in cod-liver oil) is derived from 7-dehydrocholesterol. The following chart shows the method of preparing 7-dehydrocholesterol (originated by Windaus, 1935, and improved by Buser, 1947, and by Fieser *et al.*, 1950).

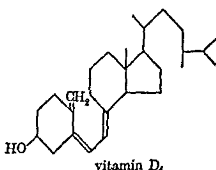


7-Dehydrocholesterol, on irradiation with ultraviolet light, gives a product that is about as active as ergocalciferol (vitamin D_2). This product was shown to be impure, and the pure active constituent was isolated as the 3:5-dinitrobenzoate (Windaus *et al.*, 1936). This vitamin D with a cholesterol side-chain is named **vitamin D_3**, and has been shown to be identical with the natural vitamin that is isolated from tunny-liver oil (Brockman, 1937). Vitamin D_3 has also been isolated from other fish-liver oils, *e.g.*, halibut. The *Chemical Society* (1951) has proposed the name **cholecalciferol** for vitamin D_3.

Irradiation of 22.23-dihydroergosterol gives a compound with antirachitic properties (Windaus *et al.*, 1937); this is known as **vitamin D_4**.

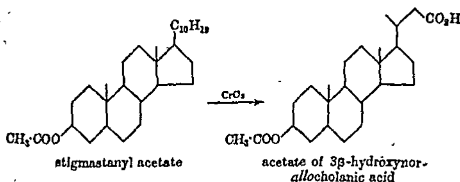


vitamin D_3



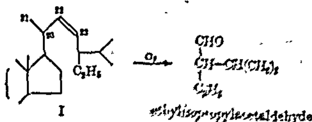
vitamin D_4

§7. **Stigmasterol,** $C_{29}H_{48}O$, m.p. 170°, is best obtained from soya bean oil. Since stigmasterol forms an acetate, etc., a hydroxyl group is therefore present. Stigmasterol also forms a tetrabromide; thus it contains two double bonds. Hydrogenation of stigmasterol produces stigmastanol, $C_{29}H_{52}O$, and since the acetate of this gives the acetate of 3*β*-hydroxynor*allo*cholanic acid on oxidation with chromium trioxide, it follows that stigmastanol differs from cholestanol only in the *nature* of the side-chain (Fernholz *et al.*, 1934;

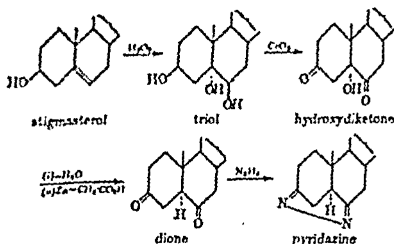


stigmastanyl acetate — acetate of 3β-hydroxynor-*allo*cholanic acid

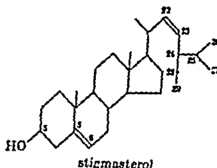
cf. ergosterol, §6). Ozonolysis of stigmasterol gives, among other products, ethyl*iso*propylacetaldehyde (Guiteras, 1933). This suggests that the side-chain is as shown in I, with a double bond at 22:23.



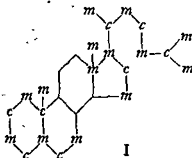
Thus the final problem is to ascertain the position of the second double bond in stigmasterol. This has been shown to be 5:6 by the method used for cholesterol (Fernholz, 1934). Stigmasterol, on hydroxylation with hydrogen peroxide in acetic acid, gives a triol which, on oxidation with chromium trioxide, forms a hydroxydiketone. This, on dehydration followed by reduction, forms a dione which combines with hydrazine to form a pyridazine derivative. These reactions can be explained as follows (cf. cholesterol, §3 b):



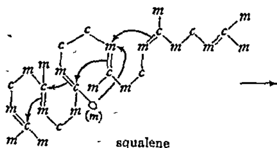
This position for the nuclear double bond is supported by other evidence; thus stigmasterol is:

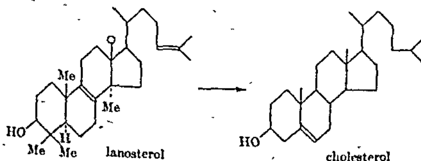


§7a. **Biosynthesis of sterols.** *It has long been known that animals can synthesise cholesterol, and work by Bloch et al. (1952), who used acetic acid labelled isotopically (with* ^{14}C*) in the methyl group (m) and in the carboxyl group (c), has established that the distribution of the carbon atoms is as shown in I. Thus acetic*



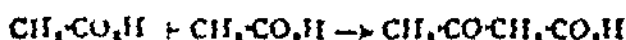
acid can be regarded as the fundamental unit, and there is evidence to show that acetoacetic acid is an intermediate. The sequence of events and the nature of other intermediates are not known with any degree of certainty. However, as we have seen in §32a. VIII, a scheme has been proposed whereby various terpenes might be synthesised in the living organism. Woodward *et al.* (1953) have suggested that the biosynthesis of cholesterol occurs *via* squalene as the intermediate. This compound contains two farnesyl chains joined end to end, thereby giving rise to the central portion being united tail to tail. The proposal is that squalene ring-closes to form lanosterol, this stage involving the shift of a methyl group (labelled 0), and then lanosterol loses three carbon atoms to form cholesterol.

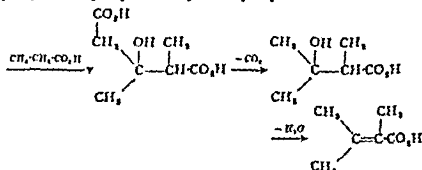




Comparison of this cholesterol structure with I shows that the distribution of the labelled carbon atoms is the same. This biosynthesis is supported by the fact that squalene, which was synthesised in rats fed with labelled acetic acid, when fed to other rats, gave cholesterol containing ^{14}C (Bloch *et al.*, 1952).

It has been suggested that the extra methyl group in the side-chain of ergosterol could arise from propionic acid (or its equivalent; *cf.* senecioic acid, §32a. VIII):

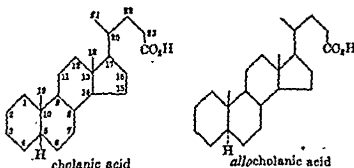




In stigmasterol, the extra methyl group presumably could arise from butyric acid (or its equivalent).

BILE ACIDS

§8. Introduction. The bile acids occur in bile (a secretion of the liver which is stored in the gall-bladder) of most animals, combined as amides with either glycine ($NH_2{\cdot}CH_2{\cdot}CO_2H$) or taurine ($NH_2{\cdot}CH_2{\cdot}CH_2{\cdot}SO_3H$), *e.g.*, glycocholic acid (= glycine + cholic acid), taurocholic acid (= taurine + cholic acid). The bile acids are present as sodium salts, and they function as emulsifying agents in the intestinal tract, *e.g.*, fats, which are insoluble in



cholanic acid *allo*cholanic acid

water, are rendered "soluble", and so may be absorbed in the intestine.

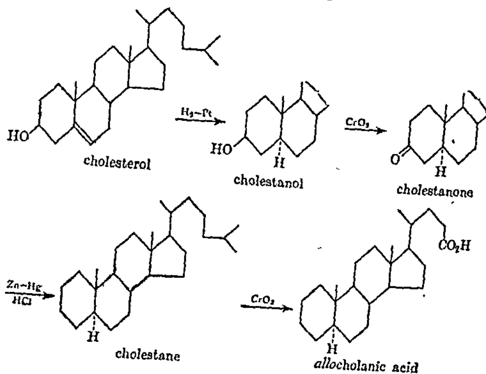
The bile acids are hydroxy derivatives of either cholanic acid or *allo*cholanic acid (but see §10). Dehydration of a bile acid by heating in a vacuum, followed by catalytic reduction, gives either cholanic or *allo*cholanic acid.

About twelve natural bile acids have been characterised, and a number of others are synthetic. The positions of the hydroxyl groups are any of the following: 3, 6, 7, 11, 12 and 23, and in almost all of the natural bile acids the configurations of the hydroxyl groups are α (see §4b). Some of the more important natural bile acids are:

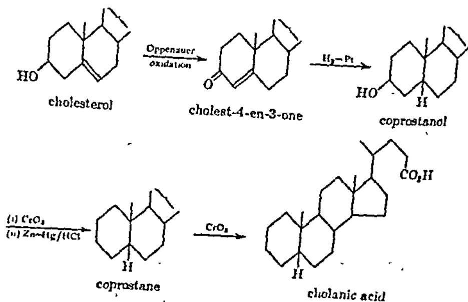
Name	*M.p.*	*Hydroxyl groups*	*Source*
Cholic acid	195°	3α:7α.12α	Man, ox
Deoxycholic acid	172°	3α.12α	Man, ox
Lithocholic acid	186°	3α	Man, ox
Chenodeoxycholic acid	140°	3α 7α	Man, ox, hen
Hyodeoxycholic acid	197°	3α 6α	Pig

§9. **The structures of cholanic acid and *allo*cholanic acid.** These acids may be derived from coprostane and cholestane, respectively, as follows (*cf.* §4c). At the same time, these reactions show the relationship between the bile acids and the sterols (Windaus, 1919).

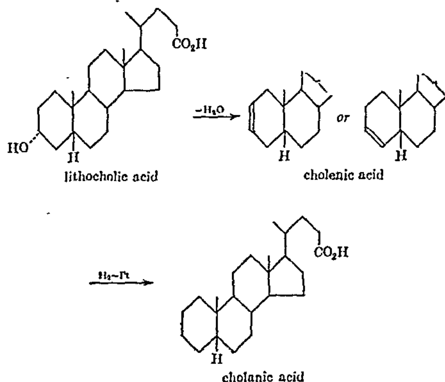
*Allo*cholanic acid.



Cholanic acid.

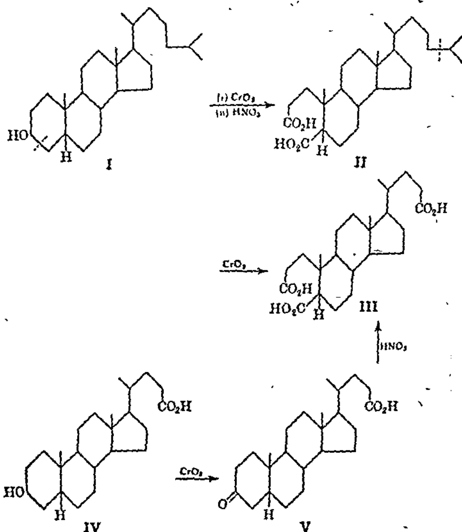


§10. Structure of the bile acids. Since all the bile acids can be converted into either of the cholanic acids, the former are therefore hydroxy derivatives of the latter, *e.g.*, lithocholic acid can be converted into cholanic acid as follows:



According to Fieser *et al.* (1955), cholenic acid is a mixture of the two compounds shown, the chol-3-enic acid being the main constituent.

The positions of the hydroxyl groups in the bile acids have been determined by means of oxidative degradation, *e.g.*, the position of the hydroxyl group in lithocholic acid is shown to be at 3 as follows. Cholesterol can be converted into coprostanol I (see, *e.g.*, §9) which, on oxidation with chromium trioxide, forms a ketone and this, when oxidised with nitric acid, gives a dicarboxylic acid, II. II, on further oxidation with nitric acid, produces the tricarboxylic acid, lithobilianic acid, III. Lithocholic acid, IV, on oxidation with chromium trioxide, forms dehydrolithocholic acid, V, and this, when oxidised with nitric acid, forms III. It therefore follows that the hydroxyl group in lithocholic acid is probably in the same position as in coprostanol, *viz.*, position 3. Thus:

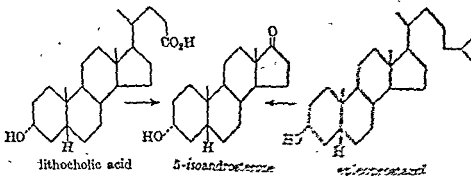


The above evidence is not conclusive, since had the hydroxyl group in lithocholic acid been at position 4, III could still have been obtained. In practice, however, the oxidation of I produces two isomeric acids for II, one being II as shown, and the other II*a*, in which the ring A is opened between C_2 and C_3; this acid, on further oxidation, gives *iso*lithobilianic acid, III*a*. Since the oxidation of lithocholic acid, IV, also produces a mixture of the *same* two acids, III and III*a*, there can be no doubt that the hydroxyl group is at position 3.

The configuration of the hydroxyl group in lithocholic acid has been shown to be α by, *e.g.*, the oxidative degradation of the acetates of lithocholic acid and epicoprostanol to 5-*iso*androsterone (formerly known as 3α-hydroxyætiocholan-17-one). Since all of the natural

HO H I → HO_2C HO_2C H IIa → CO_2H HO_2C HO_2C H IIIa

bile acids except one ("β" hyodeoxycholic acid) can be converted into lithocholic acid, all have therefore the α-configuration for the hydroxyl group at C_3.



lithocholic acid 5-isoandrosterone epicoprostanol

The bile acids form molecular compounds with various substances. Cholic acid, in particular, forms these molecular compounds with such compounds as fatty acids, esters, alcohols, etc.; these are known as the choleic acids.

The bile acids discussed in the foregoing account are all derivatives of cholanic or allocholanic acid. There are, however, some bile acids which are not derivatives of the cholanic acid, e.g., in the bile of crocodiles there is the bile acid [illegible] coprostanic acid, $C_{27}H_{46}O_5$.

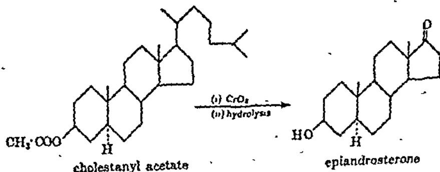
SEX HORMONES

§11. **Introduction.** Hormones are substances which are secreted by the ductless glands, and only minute amounts are necessary to produce the various physiological reactions in the body. *The sex* hormones belong to the steroid class of compounds, and are produced in the gonads (testes in the male, and ovaries in the female). Their activity appears to be controlled by the hormones that are produced in the anterior lobe of the pituitary gland. Because of this, the sex hormones are sometimes called the secondary sex hormones, and the hormones of the anterior lobe of the pituitary (which are protein in nature) are called the primary sex hormones.

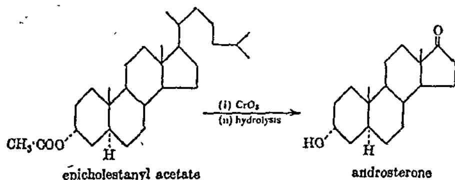
The sex hormones are of three types: the **androgens** (male hormones), the **œstrogens** (female or follicular hormones), and **progesterone** (the corpus luteum hormone). The sex hormones are responsible for the sexual processes, and for the secondary characteristics which differentiate males from females.

ANDROGENS

§12. **Androsterone,** $C_{19}H_{30}O_2$, m.p. *184–185°, is dextrorotatory.* It was first isolated by Butenandt *et al.* (1931) from male urine (about 15 mg. from 15,000 litres of urine). Androsterone behaves as a saturated compound, and since it forms mono-esters, *one* oxygen atom is present as a hydroxyl group. The functional nature of the other oxygen atom was shown to be oxo, since androsterone forms an oxime, etc. The parent hydrocarbon of androsterone, $C_{19}H_{30}O_2$, is therefore $C_{19}H_{32}$, and since this corresponds to the general formula C_nH[illegible] the molecule is tetracyclic. This led [illegible] [illegible]obably contains the steroid [illegible]e, it was thought that it is



cholestanyl acetate epiandrosterone



epicholestanyl acetate androsterone

possibly related to œstrone (§14). Butenandt (1932) therefore proposed a structure which was proved correct by Ruzicka (1934) as follows. Ruzicka oxidised cholestanyl acetate with chromium trioxide in acetic acid to **epiandrosterone**, a hydroxyketone with the structure proposed for androsterone by Butenandt. When, however, epicholestanyl acetate was oxidised, the product was androsterone. Thus the configuration of the hydroxyl group at C_3 is α and not β as Butenandt suggested. Epiandrosterone (formerly known as *iso*androsterone) has about one-eighth of the activity of androsterone.

Sondheimer *et al.* (1955) have converted epiandrosterone into androsterone, starting with epiandrosterone *p*-toluenesulphonate (*cf.* tosyl esters of sugars, §9. VII).

TsO H

AcONa
AcOH–Ac_2O

H 54%

+

AcO H 30%

$CH_2OH{\cdot}CH_2OH$

NaOH

H_2C—CH_2
O O

H

HO H

$Ph{\cdot}CO_3H$

H_2C—CH_2
O O

O H

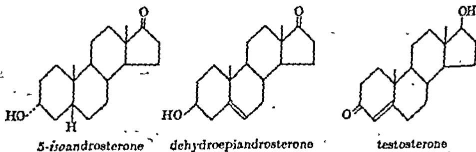
$LiAlH_4$

H_2C—CH_2
O O

HO H

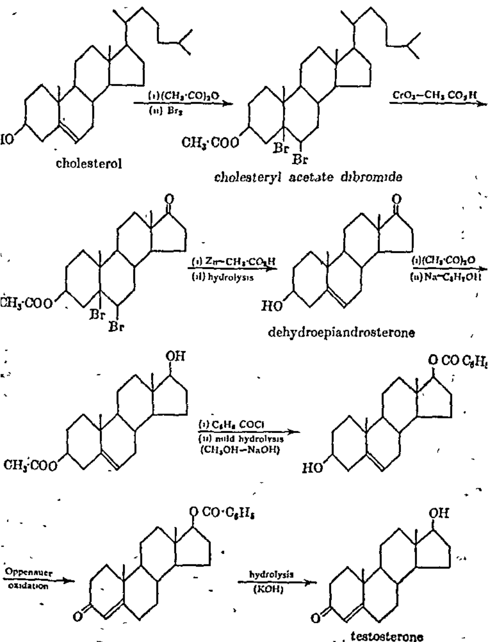
AcOH
aq

Soon after the discovery of androsterone, Butenandt *et al.* (1934) isolated two other hormones from male urine, 5-*iso*androsterone and dehydroepiandrosterone. Then Laqueur (1935) isolated the hormone testosterone from steer testes (10 mg. from 100 kg. of testes).



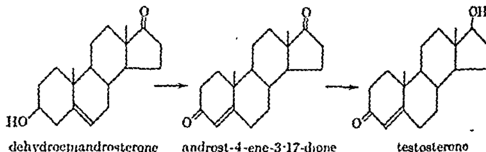
*5-iso*androsterone *dehydroepiandrosterone* *testosterone*

§13. Testosterone, $C_{19}H_{28}O_2$, m.p. 155°, is dextrorotatory. Testosterone has been produced commercially by the following method of Butenandt (1935) and Ruzicka (1935); the Oppenauer oxidation step in this method was introduced by Oppenauer (1937). This preparation of testosterone establishes the structure of this hormone.



This method has been improved by Mamoli (1938), who converted dehydroepiandrosterone into testosterone by means of micro-organisms, the first stage uses an oxidising yeast in the presence of oxygen, and the second stage a fermenting yeast.

Elisberg *et al.* (1952) have shown that sodium borohydride selectively reduces the 3-keto group in the presence of others at



dehydroepiandrosterone androst-4-ene-3:17-dione testosterone

11, 12, 17 or 20. On the other hand, Norymberski *et al.* (1954) have shown that if there is a double bond in position 4:5, then the keto group at 17 or 20 is preferentially reduced to that at 3. Thus androst-4-ene-3:17-dione is reduced to testosterone by sodium borohydride (*cf.* §3 i).

It appears that testosterone is the real male sex hormone in the body; the others are metabolic products of testosterone. The ketonic steroids are separated from the non-ketonic steroids (all from urine) by means of Girard's reagents (P and T); the ketonic compounds form *soluble* derivatives, and may be regenerated by hydrolysis (see also Vol. I). Many other hormones have also been isolated from urine.

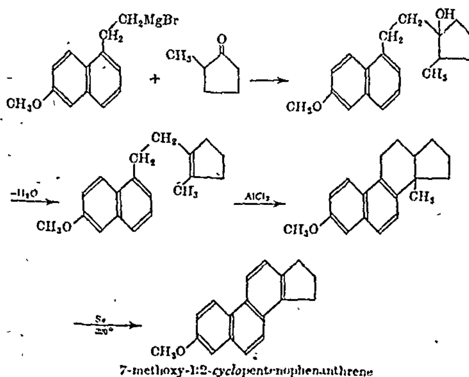
ŒSTROGENS

§14. Œstrone. It has been known for a long time that there are hormones which control the uterine cycle, but it was not until 1929 that Butenandt and Doisy independently isolated the active substance **œstrone** from the urine of pregnant women. Œstrone is the first known member of the sex hormones, and soon after its discovery two other hormones were isolated, œstriol and œstradiol.

(+)-Œstrone, m.p. 259°, has the molecular formula $C_{18}H_{22}O_2$. It behaves as a ketone (forms an oxime, etc.), and contains one hydroxyl group (it forms a monoacetate and a monomethyl ether). ... since œstrone couples ... l of ... ne;

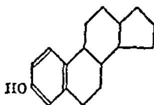
this led to the suggestion that œstrone is related to the steroids (*cf.* §1). The X-ray analysis of œstrone also indicates the presence of the steroid nucleus, and at the same time showed that the keto group and the hydroxyl group are at the opposite ends of the molecule (Bernal, 1932). On catalytic hydrogenation, œstrone forms octahydro-œstrone, $C_{18}H_{30}O_2$. This compound contains two hydroxyl groups (two hydrogen atoms are used for converting the keto group to an alcoholic group), and so six hydrogen atoms are used to saturate *three* double bonds. If these three double bonds are in one ring, *i.e.*, there is a benzenoid ring present, then the phenolic hydroxyl group can be accounted for. The presence of one benzene ring in the structure of œstrone is supported by measurements of the molecular refractivity and the ultraviolet absorption spectrum.

When the methyl ether of œstrone is subjected to the Wolff–Kishner reduction, and the product distilled with selenium, 7-methoxy-1:2-*cyclo*pentenophenanthrene is formed. The structure of this compound was established by the following synthesis (Cook *et al.*, 1934):

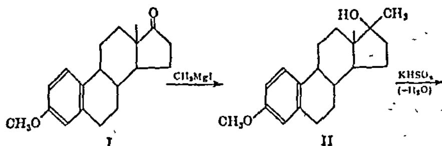


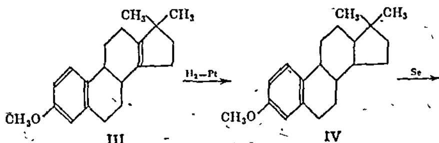
7-methoxy-1:2-*cyclo*pentenophenanthrene

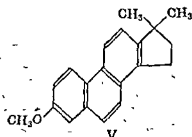
Thus the benzene ring in œstrone is ring A, and the (phenolic) hydroxyl group is at position 3; hence the skeleton of œstrone is:



Into this skeleton we must fit the keto group, and since this skeleton contains only 17 carbon atoms, another carbon atom must also be placed. The position of the keto group was shown to be at 17, and the extra carbon atom was shown to be an angular methyl group at position 13, as follows (Cook *et al.*, 1935). When the methyl ether of œstrone, I, is treated with methylmagnesium iodide, compound II is obtained. When II is dehydrated with potassium

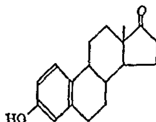






hydrogen sulphate to III, this catalytically reduced to IV, and then IV distilled with selenium, the product is 7-methoxy-3':3'-dimethyl-1:2-*cyclo*pentenophenanthrene, V. The formation of V can be explained only if there is a keto group at position 17 and an angular methyl group at position 13. It should be noted that in the following equations, the dehydration is accompanied by the migration of the angular methyl group; this assumption is based on the analogy with known examples in which this occurs.

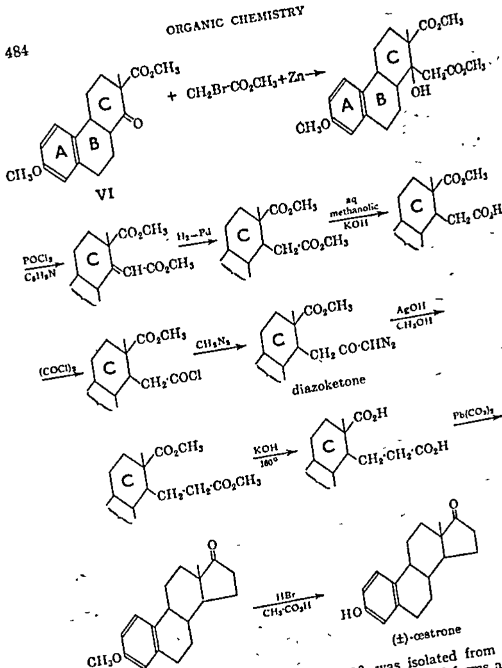
The structure of V has been confirmed by synthesis (Cook *et al.*, 1935). Thus the structure of œstrone is:



œstrone

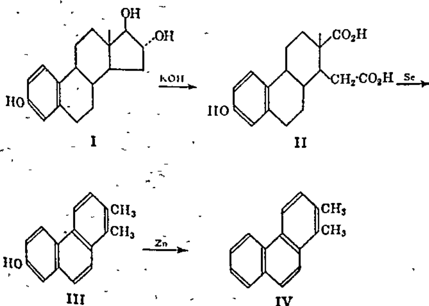
This has been confirmed by the synthesis of Anner and Miescher (1948). These authors started with the phenanthrene derivative VI, which had been prepared previously by Robinson *et al.* (1938), and by Bachmann *et al.* (1942). The first step of the Anner–Miescher synthesis involves the Reformatsky reaction, and a later one the Arndt–Eistert synthesis (see p. 484).

The stereochemical problems involved in the synthesis of œstrone are not so complicated as in cholesterol, since only four asymmetric carbon atoms are present in the hormone (*cf.* §3). VI contains 3 asymmetric carbon atoms, and so four racemates are possible. Three have been isolated by Anner and Miescher, and one of these was converted into (±)-œstrone (C/D *trans*) and the stereoisomer (C/D *cis*) as shown above. These were separated and the (±)-œstrone resolved with (−)-menthoxyacetic acid. The (+)-enantiomorph that was obtained was shown to be identical with the natural compound.



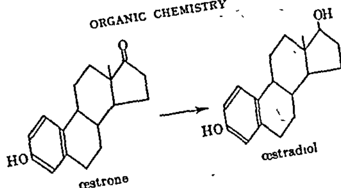
§15. Œstriol, $C_{18}H_{24}O_3$, m.p. 281°, was isolated from human pregnancy urine by Marrian (1930). Since œstriol forms a triacetate, three hydroxyl groups must be present in the molecule. One was shown to be phenolic (*cf.* œstrone), and the other two secondary alcoholic, since, on oxidation, a diketone is produced. Furthermore, X-ray analysis indicates that the two alcoholic groups are

in the *vicinal* position (*i.e.*, 1:2-). When œstriol is heated with potassium hydrogen sulphate, one molecule of water is removed and œstrone is produced. It therefore follows that œstriol has the same carbon skeleton as œstrone, and that the two alcoholic groups in œstriol are at positions 16 and 17. Structure I for œstriol fits the above facts, and is supported by the following evidence. When fused with potassium hydroxide, œstriol forms marrianolic acid, II, and this, on dehydrogenation with selenium, is converted into a hydroxydimethylphenanthrene, III, which, on distillation with zinc dust, gives a dimethylphenanthrene, IV. The structure of IV was shown to be 1:2-dimethylphenanthrene by synthesis, and since marrianolic acid forms an anhydride when heated with acetic anhydride, it therefore follows that œstriol contains a phenanthrene nucleus and a five-membered ring, the position of the latter being 1:2 (where the two methyl groups are in IV). Finally, the structure of III was shown to be 7-hydroxy-1:2-dimethylphenanthrene by synthesis (Haworth *et al.*, 1934), and so if I is the structure of œstriol, the degradation to the phenanthrene derivatives may be explained as follows:



The chemical relationship between œstrone, œstriol and œstradiol (§16) is shown by the following reactions.

(i) Œstrone may be reduced to œstradiol by catalytic hydrogenation, by aluminium *iso*propoxide (the Meerwein–Ponndorf–Verley reduction), or by lithium aluminium hydride.



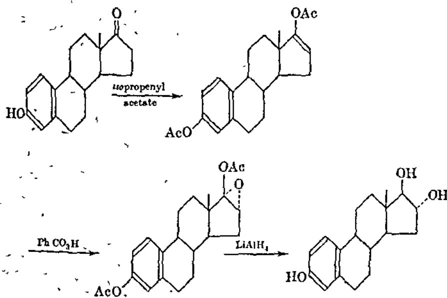
(ii) Œstriol may be converted into œstrone by the action of potassium hydrogen sulphate (see above), and œstrone may be converted into œstriol as follows (Huffman *et al.*, 1947, 1948).

CH_3O — methyl ether of œstrone $\xrightarrow[(CH_3)_3COK]{C_5H_{11}O{\cdot}NO}$ (16-oximino ketone, =NOH) $\xrightarrow[CH_3CO_2H]{Zn\ dust}$

(OH, =O; CH_3O) $\xrightarrow[(CH_3)_2CHOH]{Na}$ (OH, ·OH; CH_3O) $\xrightarrow[CH_3CO_2H]{HBr}$

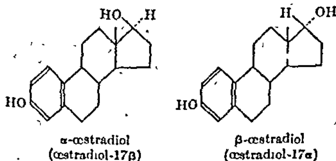
(OH, ·OH; HO)

Leeds *et al.* (1954) have converted œstrone into œstriol by a simpler method:



Œstriol is more soluble than œstrone in water, and is more potent than either œstrone or œstradiol when taken orally.

§16. **Œstradiol,** $C_{18}H_{24}O_2$. There are two stereoisomeric œstradiols, α and β; the α-isomer is much more potent than the β-.

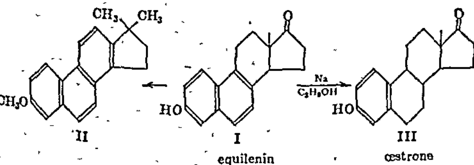


α-œstradiol
(œstradiol-17β)

β-œstradiol
(œstradiol-17α)

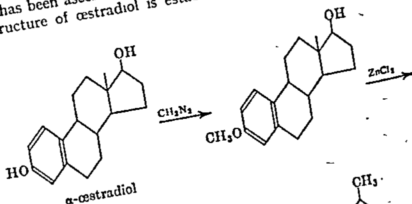
α-Œstradiol was first obtained by the reduction of œstrone (see §15), but later it was isolated from the ovaries of sows (Doisy *et al.*, 1935). When the phenolic methyl ether of œstradiol is heated with zinc chloride, a molecular rearrangement occurs, the angular methyl group migrating to the *cyclo*pentane ring D (*cf.* §2 viii. X). This compound, when dehydrogenated with selenium, produces 7-methoxy-3′-methyl-1:2-*cyclo*pentenophenanthrene, the structure of

§17. **(+)-Equilenin**, $C_{18}H_{18}O_2$, m.p. 258–259°, has been isolated from the urine of pregnant mares by Girard *et al.* (1932); it is not a very potent œstrogen. The reactions of equilenin show that a phenolic hydroxyl group and a ketonic group are present, and also that the molecule contains five double bonds (*cf.* œstrone, §14). When the methyl ether of equilenin is treated with methylmagnesium iodide, then the alcohol dehydrated, catalytically reduced and then dehydrogenated with selenium, the product is 7-methoxy-3':3'-dimethyl-1:2-*cyclo*pentenophenanthrene, II (*cf.* œstrone, §14). Thus the structure of equilenin is the same as that of œstrone, except that the former has two more double bonds than the latter (Cook *et al.*, 1935). Now the absorption spectrum of equilenin shows that it is a naphthalene derivative. Thus, since ring A in œstrone is benzenoid, it appears probable that ring B in equilenin is also benzenoid, *i.e.*, rings A and B form the naphthalene nucleus in equilenin. All the foregoing reactions of equilenin may be readily explained by assuming that I is its structure, and further evidence that has been given to support this is the claim by Marker *et al.* (1938) that equilenin may be reduced to œstrone, III, by sodium and ethanol. This reduction, however, has apparently never been substantiated (*cf.* Dauben *et al.*, 1956).

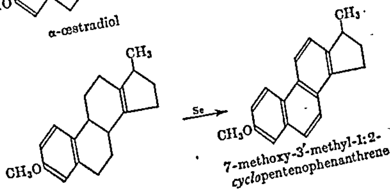


This structure of equilenin has been confirmed by synthesis. The first synthesis was by Bachmann *et al.* (1940), but was somewhat improved by Johnson *et al.* (1947). In the following chart, compound IV is synthesised by the method of Bachmann, and the rest of the synthesis is that of Johnson, who started with compound IV (Johnson's synthesis involves fewer steps than Bachmann's).

which has been ascertained by synthesis (Cook *et al.*, 1934). Thus the structure of œstradiol is established.



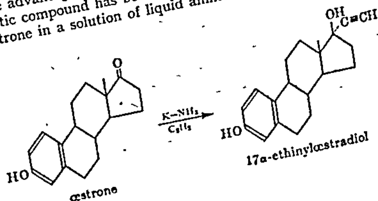
α-œstradiol



7-methoxy-3′-methyl-1:2-*cyclo*pentenophenanthrene

β-Œstradiol has been isolated from the pregnancy urine of mares (Wintersteiner *et al.*, 1938). α-Œstradiol is much more active than œstrone, whereas β-œstradiol is much less active. It appears that œstradiol is the real hormone, and that œstrone and œstriol are metabolic products.

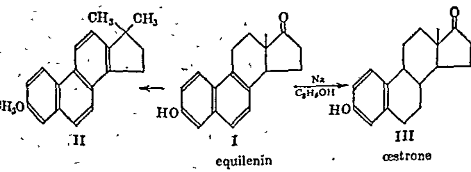
A very active synthetic œstrogen is **17α-ethinylœstradiol**, and has the advantage that it is very active when taken orally. This synthetic compound has been prepared by the action of acetylene on œstrone in a solution of liquid ammonia containing potassium.



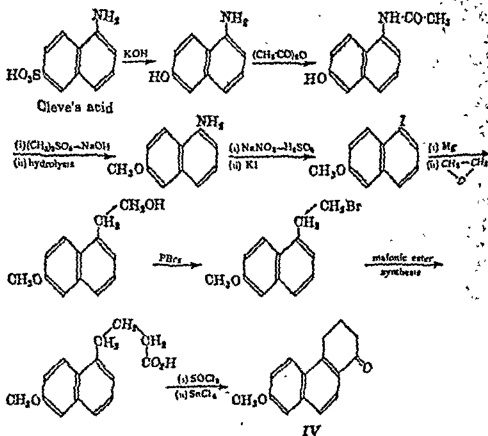
œstrone

17α-ethinylœstradiol

§17. **(+)-Equilenin,** $C_{18}H_{18}O_2$, m.p. 258–259°, has been isolated from the urine of pregnant mares by Girard *et al.* (1932); it is not a very potent œstrogen. The reactions of equilenin show that a phenolic hydroxyl group and a ketonic group are present, and also that the molecule contains five double bonds (*cf.* œstrone, §14). When the methyl ether of equilenin is treated with methylmagnesium iodide, then the alcohol dehydrated, catalytically reduced and then dehydrogenated with selenium, the product is 7-methoxy-3':3'-dimethyl-1:2-*cyclo*pentenophenanthrene, II (*cf.* œstrone, §14). Thus the structure of equilenin is the same as that of œstrone, except that the former has two more double bonds than the latter (Cook *et al.*, 1935). Now the absorption spectrum of equilenin shows that it is a naphthalene derivative. Thus, since ring A in œstrone is benzenoid, it appears probable that ring B in equilenin is also benzenoid, *i.e.*, rings A and B form the naphthalene nucleus in equilenin. All the foregoing reactions of equilenin may be readily explained by assuming that I is its structure, and further evidence that has been given to support this is the claim by Marker *et al.* (1938) that equilenin may be reduced to œstrone, III, by sodium and ethanol. This reduction, however, has apparently never been substantiated (*cf.* Dauben *et al.*, 1956).

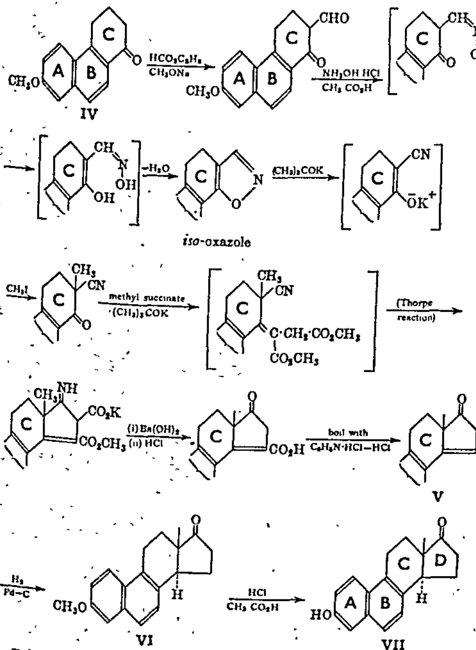


This structure of equilenin has been confirmed by synthesis. The first synthesis was by Bachmann *et al.* (1940), but was somewhat improved by Johnson *et al.* (1947). In the following chart, compound IV is synthesised by the method of Bachmann, and the rest of the synthesis is that of Johnson, who started with compound IV (Johnson's synthesis involves fewer steps than Bachmann's).



NH₂
HO₃S
Cleve's acid
KOH
HO
(CH₃CO)₂O
NH·CO·CH₃
(i) (CH₃)₂SO₄–NaOH
(ii) hydrolysis
CH₃O
(i) NaNO₂–H₂SO₄
(ii) KI
I
(i) Mg
CH₂OH
CH₂
PBr₃
CH₂Br
malonic ester synthesis
CO₂H
(i) SOCl₂
(ii) SnCl₄
O
IV

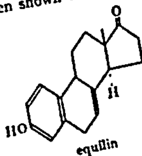
Johnson's synthesis starting from IV.



Reduction of V gives a mixture of (±)-equilenin methyl ether, VI (rings C/D *trans*), and *iso*equilenin methyl ether (rings C/D *cis*); these are separated by fractional crystallisation from acetone-methanol, the equilenin derivative being the less soluble isomer.

Product VII is (±)-equilenin, and is resolved *via* the menthoxyacetic ester. The (+)-equilenin so obtained is identical with the natural product. It should be noted here that equilenin contains only two asymmetric carbon atoms, and so the stereochemical problems involved are far simpler than those for cholesterol and œstrone.

§17a. (+)-Equilin, $C_{18}H_{20}O_2$, m.p. 238–240°, has also been isolated from the urine of pregnant mares (Girard *et al.*, 1932), and its structure has been shown to be



equilin

§18. Artificial hormones. Many compounds with œstrogenic activity, but not of steroid structure, have been prepared synthetically.

Stilbœstrol (4,4'-dihydroxydiethylstilbene) was prepared by Dodds *et al.* (1939) as follows:

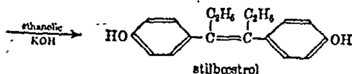
$2CH_3O \cdot C_6H_4 \cdot CHO$ (anisaldehyde) $\xrightarrow{KCN}$ $CH_3O \cdot C_6H_4 \cdot CHOH \cdot CO \cdot C_6H_4 \cdot OCH_3$ (anisoin)

$\xrightarrow{SnCl_2}$ $CH_3O \cdot C_6H_4 \cdot CH_2 \cdot CO \cdot C_6H_4 \cdot OCH_3$ (deoxyanisoin)

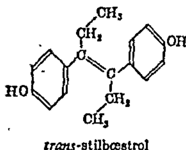
$\xrightarrow[C_2H_5I]{C_2H_5ONa}$ $CH_3O \cdot C_6H_4 \cdot CH(C_2H_5) \cdot CO \cdot C_6H_4 \cdot OCH_3$

$\xrightarrow{C_2H_5MgI}$ $CH_3O \cdot C_6H_4 \cdot CH(C_2H_5) \cdot C(C_2H_5)(OH) \cdot C_6H_4 \cdot OCH_3$

$\xrightarrow{P_2O_5}$ $CH_3O \cdot C_6H_4 \cdot C(C_2H_5){=}C(C_2H_5) \cdot C_6H_4 \cdot OCH_3$

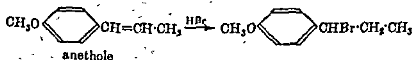


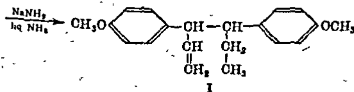
The above structure of stilbœstrol can exist in two geometrical isomeric forms; it is the *trans*-form which is the active substance, and this configuration has been confirmed by X-ray analysis (Crowfoot *et al.*, 1941).

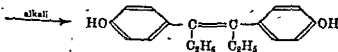


trans-stilbœstrol

Kharasch *et al.* (1943) have introduced a simpler synthesis of stilbœstrol. Anethole is treated with hydrobromic acid and the product, anethole hydrobromide, is then treated with sodamide in liquid ammonia. The resulting compound, I, gives stilbœstrol on demethylation and isomerisation in the presence of alkali. The structure of I is uncertain, but it is believed to be the one given.

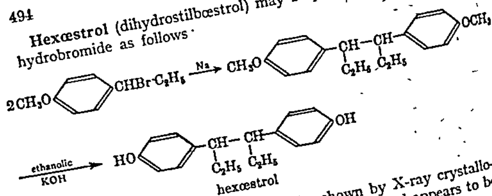






Stilbœstrol is more active than œstrone when administered subcutaneously, and it can also be given orally.

Hexœstrol (dihydrostilbœstrol) may be prepared from anethole hydrobromide as follows:



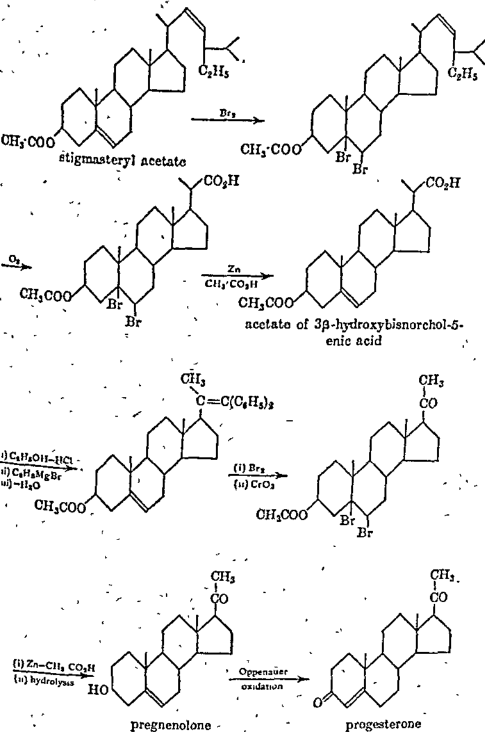
The active form is the *meso*-isomer (as shown by X-ray crystallography by Crowfoot *et al.*, 1941), and this compound appears to be the most potent of the œstrogens.

GESTOGENS

§19. Progesterone, $C_{21}H_{30}O_2$, m.p. 128°, was first isolated in a pure form by Butenandt *et al.* (1934) from the *corpora lutea* of pregnant sows.

The chemical reactions of progesterone show that there are two keto groups present, and since on catalytic reduction three molecules of hydrogen are added to form the dialcohol $C_{21}H_{36}O_2$, it therefore follows that progesterone contains one double bond (four hydrogen atoms are used to convert the two keto groups to alcoholic groups). Thus the parent hydrocarbon of progesterone is $C_{21}H_{36}$, and since this corresponds to the general formula C_nH_{2n-6}, progesterone is therefore tetracyclic. Furthermore, X-ray studies have shown that progesterone contains the steroid nucleus, and this is further supported by the fact that progesterone may be prepared from, *e.g.*, stigmasterol and cholesterol. These preparations also show the structure of progesterone, but do not provide conclusive evidence for the position of the double bond in progesterone, since the results can be interpreted equally well on the assumption that the double bond is 4:5 or 5:6. The absorption spectrum of progesterone, however, shows that it is an $\alpha\beta$-unsaturated ketone, and this suggests that the position of the double bond is 4:5 (see below). Finally, progesterone has also been synthesised from diosgenin and from pregnanediol, and the preparation from the latter, taken in conjunction with the others, definitely shows that the position of the double bond in progesterone is 4:5.

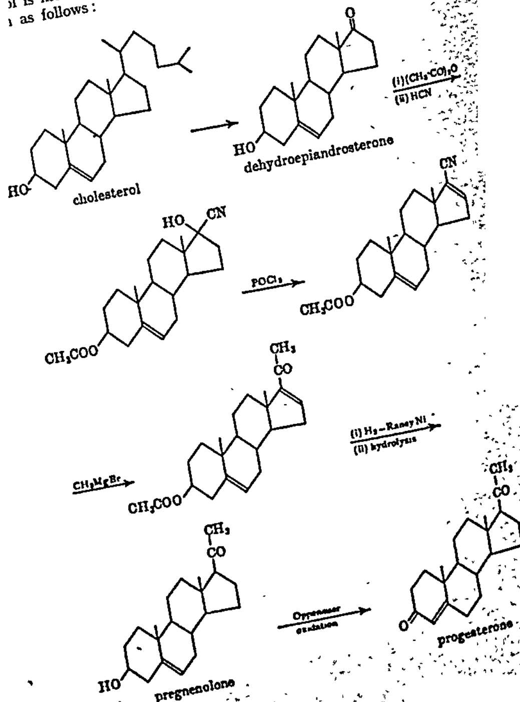
(i) *Progesterone from stigmasterol* (Butenandt *et al.*, 1934, with improvements by other workers).



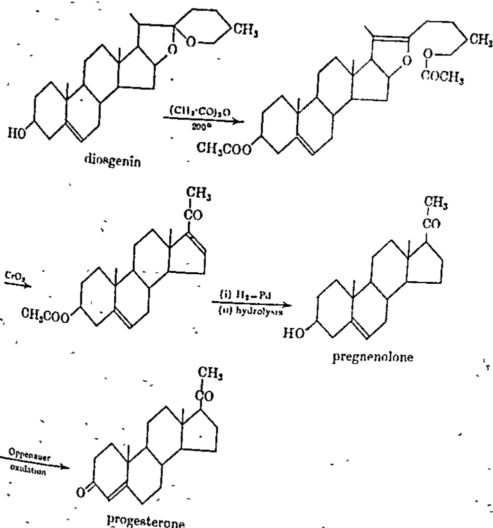
pregnenolone progesterone

Pregnenolone has also been isolated from the *corpus luteum.*

Progesterone from cholesterol (Butenandt *et al.*, 1939). Cholesterol is first converted into dehydroepiandrosterone (see §13), and then as follows:



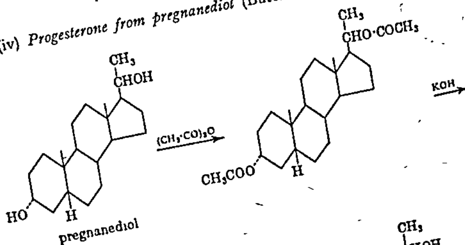
(iii) *Progesterone from diosgenin* (Marker *et al.*, 1940, 1941). Diosgenin (a sapogenin) occurs as a glycoside in the root of *Trillium erectum*.

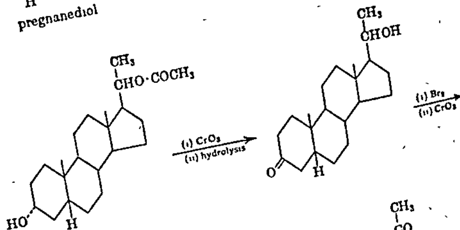


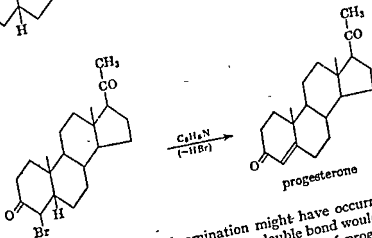
Saponins and Sapogenins. Saponins are plant glycosides, and the aglycon is known as the sapogenin (*cf* §24 VII). Saponins are very powerful emulsifiers, and derive their name from this property; they are used as detergents. There are two groups of saponins, the steroid and the triterpenoid saponins, and these two groups may be distinguished by the fact that only the former group gives Diels' hydrocarbon on distillation with selenium; the triterpenoid group gives mainly naphthalene or picene derivatives (*cf.* §1).

Digitonin is a steroid saponin; it causes hæmolysis of the red blood cells

(iv) *Progesterone from pregnanediol* (Butenandt *et al.*, 1930).

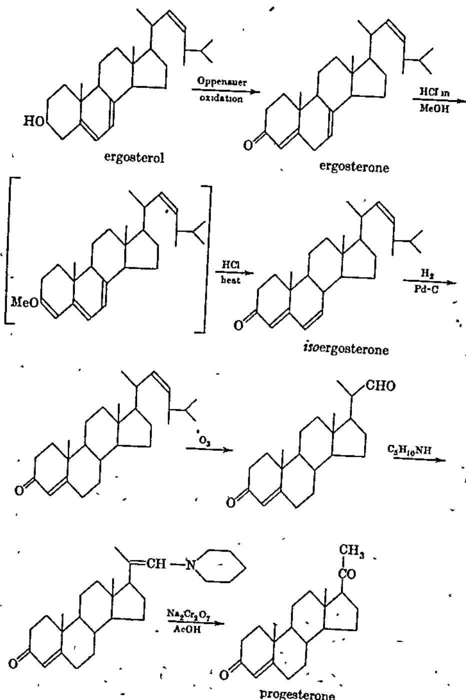




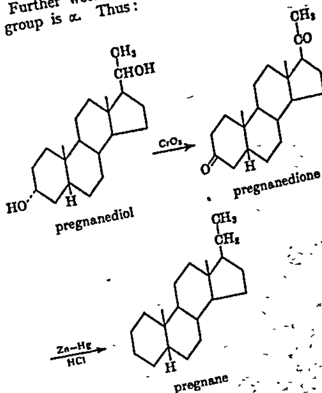


In the above reactions, bromination might have occurred in position 2; in this case the position of the double bond would have been 1:2. This is impossible, since the preparation of progesterone by methods (i) to (iii) shows that the double bond must be 4:5 or 5:6. Thus the preparation from pregnanediol proves that the double bond is 4:5.

(v) *Progesterone from ergosterol* (Shepherd *et al.*, 1955). This appears to be the most practical synthesis.



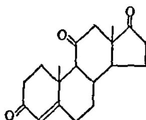
§20. **Pregnane-3α:20α-diol**, $C_{21}H_{36}O_2$, was isolated from human pregnancy urine by Marrian (1929); it is biologically inactive, and is the main metabolic product of progesterone. The functional nature of the two oxygen atoms was shown to be secondary alcoholic, and since pregnanediol is saturated, the parent hydrocarbon is $C_{21}H_{36}$, and so the molecule is tetracyclic. Pregnanediol gives the haloform reaction; thus a $CH_3 \cdot CHOH \cdot$ group is present (see Vol. I). When oxidised, pregnanediol is converted into the diketone pregnanedione and this, on the Clemmensen reduction, forms pregnane, $C_{21}H_{36}$. This is identical with 17-ethylætiocholane, a compound of known structure. Thus pregnanediol contains the steroid nucleus, and the position of the side-chain is 17. Finally, the relationship between pregnanediol and progesterone shows that the former contains one hydroxyl-group at position 3. Further work showed that the configuration of the 3-hydroxyl group is α. Thus:



ADRENAL CORTICAL HORMONES

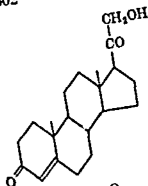
§21. Introduction. In the adrenal glands (of mammals) there are two regions, the *medulla* which produces adrenaline (see §12. XIV), and the *cortex* which produces steroid hormones. The production of these adreno-cortical hormones is controlled by the hormone produced in the anterior lobe of the pituitary, the so-called adrenocorticotrophic hormone, ACTH.

§22. Adrenal cortical hormones. About 28 steroids have been isolated from the extract of the adrenal cortex, and their structures have been elucidated mainly by Kendall *et al.* (1935), Wintersteiner (1935–) and Reichstein *et al.* (1936–). Only six of these 28 compounds are physiologically active, fourteen are inactive and are produced by the reduction of the active hormones, and the remaining six are œstrone, progesterone, 17α-hydroxyprogesterone and adrenosterone, and two other compounds that are apparently produced by oxidation during the isolation of the hormones from the cortical extract. Adrenosterone is as shown, and possesses androgenic activity.

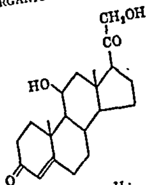


adrenosterone

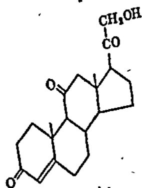
The six active compounds are as follows (they have been designated by letters as well as named systematically).



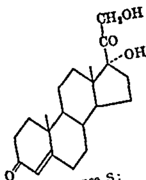
Substance Q, 11-Deoxycorticosterone; 21-Hydroxyprogesterone



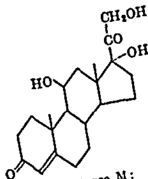
Substance H; Corticosterone; 11:21-Dihydroxyprogesterone



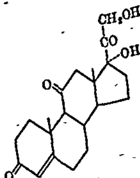
Compound A; 11-Dehydrocorticosterone 21-Hydroxy-11-ketoprogesterone



Substance S; 11-Deoxy-17-hydroxycorticosterone



Substance M; 17-Hydroxycorticosterone



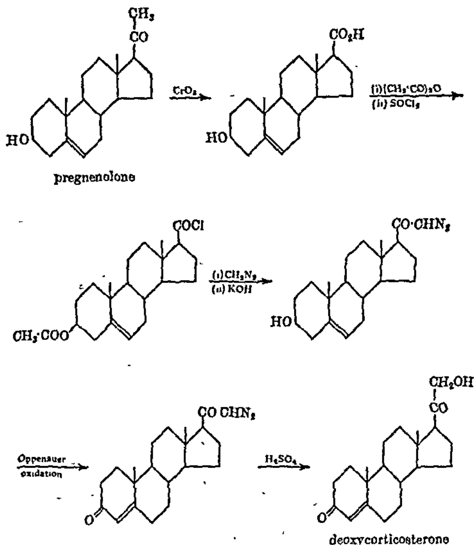
Substance F; Compound E; 11-Dehydro-17-hydroxycorticosterone; cortisone

Owing to the presence of the α-hydroxyketone group, the adrenal cortical hormones are strong reducing agents. The hydroxyl group at position 21 behaves in the usual way, but the 11-keto group does not form an oxime or a phenylhydrazone. The 11-keto group is resistant to catalytic reduction in neutral solution, but can be reduced in acid solution; it is readily reduced to a hydroxyl group by lithium aluminium hydride, and to a methylene group by the Clemmensen reduction.

The keto-hormones are separated from non-keto compounds by means of Girard's reagents P and T (see Vol. I).

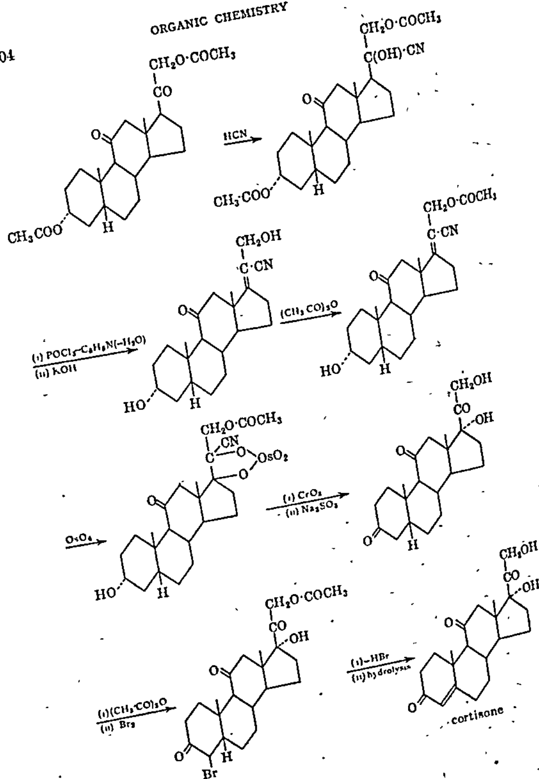
The structures of the cortical hormones have been elucidated by degradation and by partial syntheses from sterols of known structure, *e.g.*, deoxycorticosterone from stigmasterol (Reichstein *et al.*,

1937, 1940). The first step is the conversion of stigmasterol to pregnenolone (see §19 i).



A very interesting point about the above synthesis is the unusual stability of the diazoketone.

Cortisone (Substance F, Compound E) has been used for the treatment of rheumatoid arthritis and rheumatic fever. Many partial syntheses are known; and there is also a total synthesis; *e.g.*, the following partial synthesis starts from 3α:21-diacetoxypregnane-11:20-dione (Sarett, 1948).

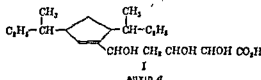


CH2O·COCH3
CO
CH3COO
H
HCN
CH2O·COCH3
C(OH)·CN
CH3·COO
H
CH2OH
C·CN
(i) POCl3–C5H5N(–H2O)
(ii) KOH
HO
H
(CH3CO)2O
CH2O·COCH3
C·CN
HO
H
CH2O·COCH3
CN
C
O
OsO2
O
OsO4
HO
H
(i) CrO3
(ii) Na2SO3
CH2OH
CO
OH
H
CH2O·COCH3
CO
OH
(i)(CH3CO)2O
(ii) Br2
H
Br
(i)–HBr
(ii) hydrolysis
CH2OH
CO
OH
cortisone

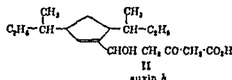
AUXINS

§23. [illegible]

isolated an active compound from human urine, and they named it **auxin** *a* and showed that its structure is I. Soon afterwards, Kögl *et al.* isolated **auxin** *b* (II) from maize germ oil.



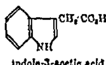
I
auxin *a*



II
auxin *b*

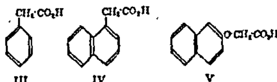
The name **auxin** is now taken as the *generic* name for the plant hormones. Auxins generally occur in the plant kingdom, but are also present in urine, etc.

Further work by Kögl *et al.* (1934) led to the isolation from urine of another growth-promoting substance which the authors named "hetero-auxin", and subsequently showed that this compound is indole-3-acetic acid.



indole-3-acetic acid

The discovery that indole-3-acetic acid had plant growth-promoting properties led to the examination of compounds of related structure, and it was soon found that various derivatives of indole-3-acetic acid are also active; it was also found that a number of arylacetic acids and aryloxyacetic acids are active, *e.g.*, phenylacetic acid, III, 1-naphthaleneacetic acid, IV, and 2-naphthoxyacetic acid, V.



Recent work has suggested that indole-3-acetic acid is the natural plant hormone, and not auxins *a* and *b*. In fact, there now appears to be some doubt as to the existence of auxin *a* (*auxentriolic acid*) and

ı b (*auxenolonic acid*); neither of these compounds has been ted since Kogl obtained them.

he relation between chemical structure and growth-promoting pro- ies has still to be solved, but nevertheless much progress has been de in this direction. Koeplı *et al.* (1938) believe that a plant hor- ne must have a ring structure containing at least one double bond, d a side-chain containing a carboxyl group (or a group capable of ing converted into a carboxyl group) removed from the ring by at ast one carbon atom (*cf.* compounds I–V). These requirements, owever, have been modified by Veldestra (1944–).

READING REFERENCES

Fieser and Fieser, *Natural Products Related to Phenanthrene*, Reinhold (1949, 3rd ed.).

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). Ch. 19. The Steroids.

Rodd (Ed.), *Chemistry of Carbon Compounds*, Elsevier. Vol. IIB (1953). Ch. 17. Sterols and Bile Acids. Ch. 18. Sex Hormones; Adrenocortical Hormones.

Stewart and Graham, *Recent Advances in Organic Chemistry*, Longmans, Green. Vol. III (1948, 7th ed.). Ch. I. The Bile Acids and Sterols. Ch. III. The Hormones.

Hormones, The Pharmaceutical Press (1951).

Vitamins and Hormones, Academic Press (Vol. I, 1943–).

Sobotka, *The Chemistry of the Sterids*, Williams and Wilkins (1938).

Cook (Ed.), *Progress in Organic Chemistry*, Butterworth. Vol. II (1953). Ch. 4. The Partial Synthesis of Cortisone and Related Compounds from Accessible Steroids. Ch. 5. The Relationship of Natural Steroids to Carcinogenic Aromatic Compounds. Vol. III (1955). Ch. 1. Total Synthesis of Steroids.

Klyne, *The Chemistry of the Steroids*, Methuen (1957).

Editorial Report on Nomenclature, *J.C.S.*, 1951, 3526. Appendix C. Steroid Nomenclature.

Heilbron *et al.*, The Structure of the Sterols and Bile Acids, *J.C.S.*, 1933, 626.

Spring, Recent Advances in the Chemistry of the Steroids, *J.C.S.*, 1950, 3352.

Ann. Reports (*Chem. Soc.*); Steroids (Total Synthesis), 1950, 47, 205; 1951, 48, 200; 1952, 49, 190; 1953, 50, 216; 1954, 51, 222; 1955, 52, 207; 1956, 53, 216.

Woodward *et al.*, The Total Synthesis of Steroids, *J. Amer. Chem. Soc.*, 1952, 74, 4223.

Robinson *et al.*, (i) Synthesis of a Tricyclic Degradation Product of Cholesterol, *J.C.S.*, 1949, 1855. (ii) Completion of the Syntheses of Androgenic Hormones and of the Cholesterol Group of Sterols, *J.C.S.*, 1953, 361.

Dodds, Synthetic Œstrogens, *J. Pharm. Pharmacol.*, 1949, 1, 137.

Hems, The Chemistry of Cortisone, *J. Pharm. Pharmacol.*, 1953, 5, 409.

Wicker, The Mechanism of Catalytic Hydrogenation of Cyclic Compounds, *J.C.S.*, 1956, 2165.

Popják, Chemistry, Biochemistry and Isotopic Tracer Technique, Royal Institute of Chemistry Monograph, No. 2 (1955).
Skoog (Ed.), *Plant Growth Substances*, University of Wisconsin (1951).
Pincus and Thimann (Ed.), *The Hormones*, Academic Press. Vol. I (1948). Plant Growth Hormones (p. 5).
Audus, *Plant Growth Substances*, Leonard Hill Ltd. (1953).

CHAPTER XII

HETEROCYCLIC COMPOUNDS CONTAINING TWO OR MORE HETERO-ATOMS

§1. Nomenclature. (i) When the heterocyclic compound contains two or more hetero-atoms, the starting point for numbering is the hetero-atom of as high a group in the periodic table and as low an atomic number in that group. Thus the order of naming will be O, S, Se, N, P, As, Sb, Si, Sn, Pb, Hg.

(ii) With the atom of the preferred kind as number 1, the ring is numbered in such a way that the hetero-atoms are given the lowest numbers possible.

(iii) Of two or more numberings conforming to rules (i) and (ii), the one that is chosen is that which assigns low numbers more nearly in the order of precedence established by rule (i).

(iv) Of two or more numberings conforming to rules (i)–(iii), the one that is chosen is that which gives hydrogen atoms the lowest numbers possible.

(v) When a heterocyclic compound containing at least one nitrogen atom does not end in *ine* and gives *basic* compounds on progressive hydrogenation, the latter derivatives will be indicated by the successive endings *ine, idine*; *e.g.*, pyrazole, pyrazoline, pyrazolidine.

The hetero-atoms in heterocyclic compounds are indicated by prefixes, *e.g.*, O by **oxa**, S by **thia**, N by **aza**.

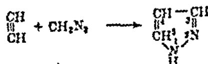
AZOLES

Azole is the suffix used for five-membered rings containing two or more hetero-atoms, at least one of which is nitrogen.

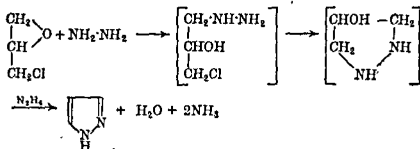
PYRAZOLE GROUP

§2. Pyrazole. Pyrazole may be synthesised in a number of ways, some of the more convenient methods being the following:

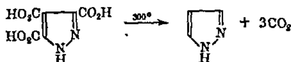
(i) By passing acetylene into a cold ethereal solution of diazomethane (von Pechmann, 1898).



(ii) By heating epichlorohydrin with hydrazine in the presence of zinc chloride (Balbiano, 1890).

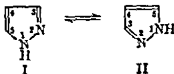


(iii) By the decarboxylation of various pyrazolecarboxylic acids, *e.g.*, by heating pyrazole-3:4:5-tricarboxylic acid (see also §2a, ii).



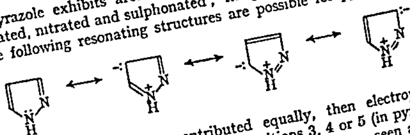
(iv) Jones (1949) has shown that pyrazole may be conveniently prepared by the condensation of 1:1:3·3-tetraethoxypropane, $(C_2H_5O)_2CH{\cdot}CH_2{\cdot}CH(OC_2H_5)_2$, with hydrazine dihydrochloride.

Properties of pyrazole. Pyrazole is a colourless solid, m.p. 70°. It is a tautomeric substance; the existence of tautomerism cannot be demonstrated in pyrazole itself, but it can be inferred by the consideration of pyrazole derivatives. If pyrazole is tautomeric, then the positions 3 and 5 will be identical; if pyrazole is not tautomeric, then these positions are different. Now Knorr *et al.* (1893) showed that on oxidation, both 3-methyl-1-phenylpyrazole and 5-methyl-1-phenylpyrazole gave the *same* product, *viz.*, methylpyrazole. Thus positions 3 and 5 must be equivalent in pyrazole,

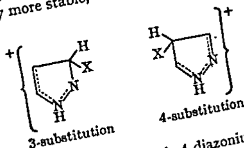


and this can only be explained by assuming that pyrazole is tautomeric (I and II). It therefore follows that in pyrazole there can only be *two* carbon-alkyl derivatives, 3- (or 5-) and 4-. If, however, the imino hydrogen is replaced by an alkyl or aryl group, then *three* carbon-alkyl derivatives are possible, 3, 4 and 5, since tautomerism is now impossible, and so positions 3 and 5 are no longer equivalent.

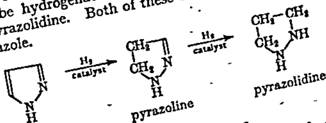
Pyrazole exhibits aromatic properties, *e.g.*, it is readily halogenated, nitrated and sulphonated; the group enters at position 4. The following resonating structures are possible for pyrazole.



If these structures are contributed equally, then electrophilic attack should occur equally well at positions 3, 4 or 5 (in pyrazole itself, positions 3 and 5 are equivalent). As we have seen above, electrophilic attack occurs exclusively at position 4. The reason for this is not certain. Possibly the resonating structures are not contributed equally (as was assumed). On the other hand, Dewar (1949) has suggested that substitution occurs in the 4-position because the transition state for 4-substitution is more symmetrical, and consequently more stable, than the transition state for 3- (or 5-) substitution.

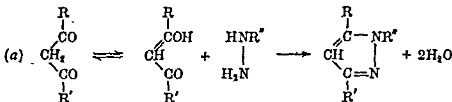


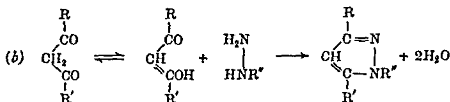
It is interesting to note that pyrazole-4-diazonium salts are stable to boiling water. Pyrazole is feebly basic, and forms salts with inorganic acids; the imino hydrogen may be replaced by an acyl group. Pyrazole is very resistant to oxidising and reducing agents, but may be hydrogenated catalytically, first to pyrazoline, and then to pyrazolidine. Both of these compounds are stronger bases than pyrazole.



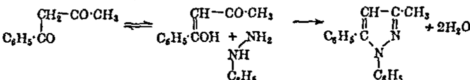
§2a. Synthesis of pyrazole derivatives.

(i) A very important method for preparing pyrazole derivatives is by the reaction between β-diketones (or β-ketoaldehydes) and hydrazines (Knorr *et al.*, 1883).

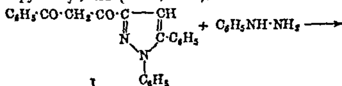


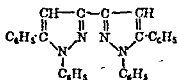


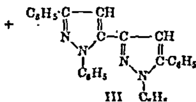
Thus, according to the above, a mixture of isomeric pyrazoles will be produced. In actual practice, the main product depends on the nature of R and R', and it appears that if R is an alkyl group and R' an aryl group, then only *one* product is obtained, *e.g.*, benzoylacetone and phenylhydrazine form only 3-methyl-1:5-diphenylpyrazole (Drumm, 1931).



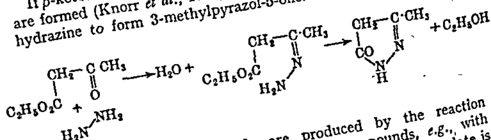
Two different enolic forms are possible for benzoylacetone, but it appears that only the enol $C_6H_5{\cdot}C(OH) = CH{\cdot}CO{\cdot}CH_3$ is present, and this is possibly due to the fact that in this form extended conjugation with the benzene ring can occur and so stabilise this form. On the other hand, if both R and R' are aryl groups, then two isomers may be isolated, *e.g.*, 3-α-benzoylacetyl-1:5-diphenylpyrazole, I, reacts with phenylhydrazine to produce a mixture of 1:1':5:5'-tetraphenyl-3.3'-dipyrazolyl, II, and 1:1'.3':5-tetraphenyl-3·5'-dipyrazolyl, III (Finar, 1955).



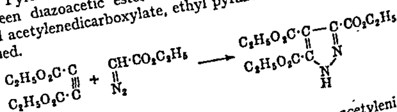




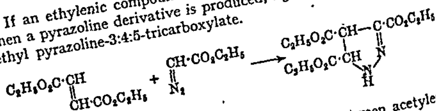
If β-ketoesters are used instead of β-diketones, then 5-pyrazolones are formed (Knorr *et al.*, 1883), *e.g.*, ethyl acetoacetate reacts with hydrazine to form 3-methylpyrazol-5-one.



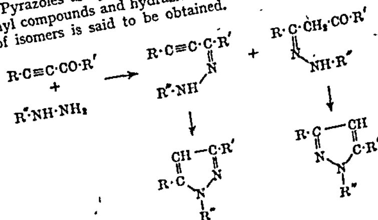
(ii) Pyrazolecarboxylic acids are produced by the reaction between diazoacetic ester and acetylenic compounds, *e.g.*, with ethyl acetylenedicarboxylate, ethyl pyrazole-3:4:5-tricarboxylate is formed.



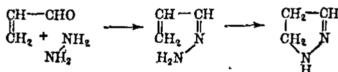
If an ethylenic compound is used instead of an acetylenic one, then a pyrazoline derivative is produced, *e.g.*, ethyl fumarate gives ethyl pyrazoline-3:4:5-tricarboxylate.



(iii) Pyrazoles are produced by the reaction between acetylenic carbonyl compounds and hydrazines (Moureu *et al.*, 1903); a mixture of isomers is said to be obtained.

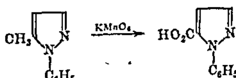


(iv) Pyrazolines are obtained by the condensation of α,β-unsaturated ketones or aldehydes with hydrazines, *e.g.*, acraldehyde and hydrazine give pyrazoline.

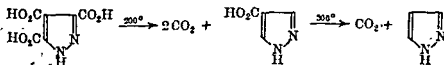


Pyrazolines may be oxidised to pyrazoles by bromine or mercuric oxide.

Properties of the pyrazole derivatives. Pyrazoles with substituent methyl groups may be oxidised by potassium permanganate to the corresponding pyrazolecarboxylic acids, *e.g.*,



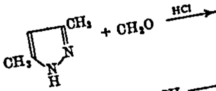
Pyrazole-3- and 5-carboxylic acids are readily decarboxylated by heating above their melting points; the pyrazole-4-carboxylic acids are more stable, but can nevertheless be decarboxylated at elevated temperatures, *e.g.*,

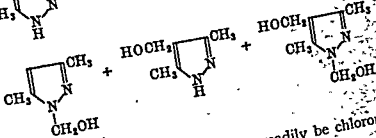


Although pyrazole itself is not reduced by sodium and ethanol, *N*-phenyl substituted pyrazoles are readily reduced to the corresponding pyrazolines, *e.g.*,



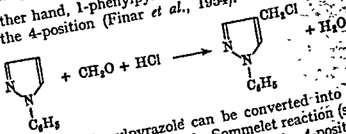
1-Unsubstituted pyrazoles apparently cannot be chloromethylated; carbinols are produced, *e.g.* (Dvoretzky *et al.*, 1950);





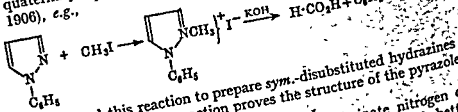
(main product)

...n the other hand, 1-phenylpyrazole can readily be chloromethyl...ed in the 4-position (Finar *et al.*, 1954).



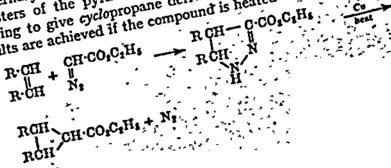
4-Chloromethyl-1-phenylpyrazole can be converted into 1-phenylpyrazole-4-aldehyde by means of the Sommelet reaction (see Vol. I). 1-Phenylpyrazole can also be mercurated in the 4-position (Finar *et al.*, 1954).

When boiled with concentrated aqueous potassium hydroxide, quaternary pyrazoles are converted into hydrazines (Knorr *et al.*, 1906), *e.g.*,

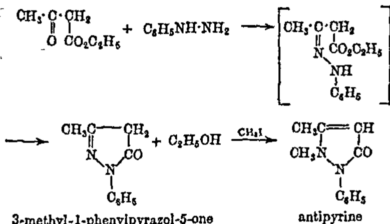


Knorr used this reaction to prepare *sym.*-disubstituted hydrazines; at the same time, this reaction proves the structure of the pyrazole-quaternary salts.

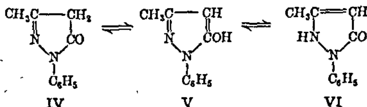
Esters of the pyrazolinecarboxylic acids eliminate nitrogen on heating to give *cyclo*propane derivatives; sometimes much better results are achieved if the compound is heated with copper powder.



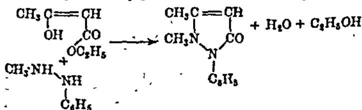
Antipyrine (2:3-dimethyl-1-phenylpyrazol-5-one), m.p. 127°, is very much used in medicine as a febrifuge. It is prepared industrially by condensing ethyl acetoacetate with phenylhydrazine, and methylating the product, 3-methyl-1-phenylpyrazole-5-one, with methyl iodide in alkaline ethanolic solution, or with methyl sulphate in the presence of sodium hydroxide.



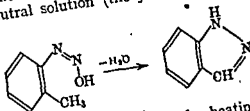
At first sight one might have expected to obtain the *O*-methyl or the 4-methyl derivative, since the tautomeric forms IV (keto) and V (enol) are theoretically possible. Methylation of 3-methyl-1-phenylpyrazole-5-one with diazomethane results in the formation of



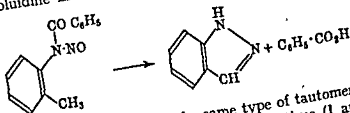
the *O*-methyl derivative (this is also produced in a small amount when methyl iodide is used as the methylating reagent). This raised some doubts as to the structure of antipyrine, since for its formation, the tautomeric form VI must also be postulated. The structure of antipyrine was shown to be that given above by its synthesis from *sym*.-methylphenylhydrazine and ethyl acetoacetate.



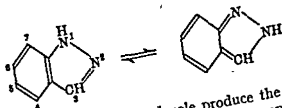
§2b. Indazoles (benzopyrazoles). Indazole may be prepared by the removal of a molecule of water from *o*-toluenediazohydroxide in neutral solution (the yield is very poor).



Indazole may conveniently be prepared by heating *o*-*N*-nitroso-*N*-benzoyltoluidine in benzene solution.



Indazole, m.p. 146°, exhibits the same type of tautomerism that exists in pyrazole, since *two* series of *N*-derivatives (1 and 2) are known:



Nitration and sulphonation of *indazole* produce the 5-substitution product; bromination gives the 3:5-dibromo compound.

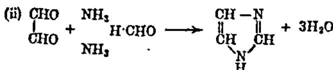
GLYOXALINE GROUP

This group of compounds is also known as the *iminazoles* or the *imidazoles*.

§3. Glyoxaline (iminazole, imidazole) is isomeric with pyrazole, and occurs in the purine nucleus and in the amino-acid histidine; 4-aminoglyoxaline-5-carboxamide occurs naturally as a riboside (or ribotide).

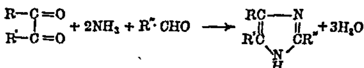
Glyoxaline may be prepared by the action of ammonia on glyoxal. The mechanism of this reaction is uncertain, but one suggestion is that one molecule of glyoxal breaks down into formic acid and formaldehyde, and then the latter reacts as follows:



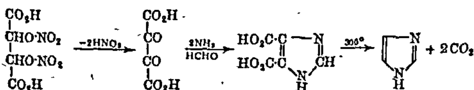


A certain amount of support for this mechanism is given by the fact that glyoxaline may be prepared directly from glyoxal, ammonia and formaldehyde.

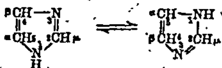
A general method for preparing glyoxalines is by the reaction between an α-dicarbonyl compound, ammonia and an aldehyde (Radziszewsky, 1882).



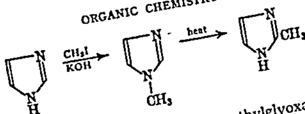
Glyoxaline itself is best prepared by the action of ammonia on a mixture of formaldehyde and tartaric acid dinitrate ("dinitrotartaric acid"), and then heating the dicarboxylic acid thereby produced.



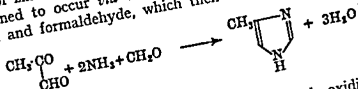
Glyoxaline, m.p. 90°, is a weak base, but it is more basic than pyrazole. Glyoxaline is a tautomeric substance, since positions 4 and 5 are equivalent (positions 5, 4 and 2 have also been designated α, β and μ, respectively).



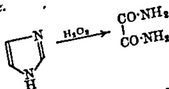
Methyl iodide attacks glyoxaline in potassium hydroxide solution to form 1-methylglyoxaline which, when strongly heated, isomerises to 2-methylglyoxaline (*cf.* the Hofmann rearrangement; see Vol. I).



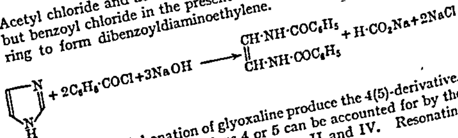
An interesting method of preparing 4(5)-methylglyoxaline is by the action of zinc hydroxide and ammonia on glucose; the reaction is assumed to occur *via* the breakdown of glucose into methylglyoxal and formaldehyde, which then react as follows:



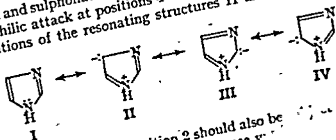
The glyoxaline ring is extremely stable towards oxidising and reducing agents; hydrogen peroxide, however, readily opens the ring to form oxamide.



Acetyl chloride and acetic anhydride have no action on glyoxaline, but benzoyl chloride in the presence of sodium hydroxide *opens* the ring to form dibenzoyldiaminoethylene.

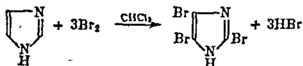


Nitration and sulphonation of glyoxaline produce the 4(5)-derivative. Electrophilic attack at positions 4 or 5 can be accounted for by the contributions of the resonating structures II and IV. Resonating



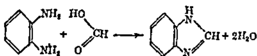
structure III shows that position 2 should also be [illegible] to electrophilic attack. This is found to be the case [illegible] *e.g.*,

bromine reacts with glyoxaline in chloroform solution to give 2:4:5-tribromoglyoxaline.



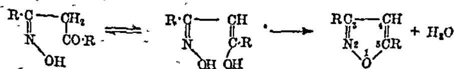
Glyoxaline couples with diazonium salts in the 2-position, but *N*-alkylglyoxalines do not couple at all.

§3a. **Benzoglyoxalines (benziminazoles).** These are readily formed by heating *o*-phenylenediamines with carboxylic acids, *e.g.*, benziminazole itself (m.p. 170°) is produced by heating *o*-phenylenediamine with 90 per cent. formic acid.

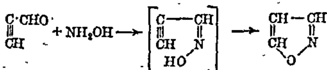


OXAZOLE GROUP

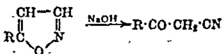
§4. ***iso*-Oxazoles.** *iso*-Oxazoles are formed by the dehydration of the monoximes of β-diketones or β-ketoaldehydes.



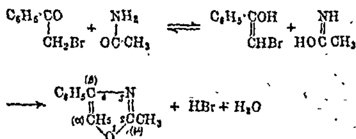
iso-Oxazole itself may be prepared by the action of hydroxylamine on propargylaldehyde.



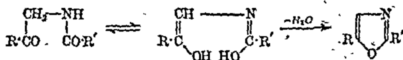
iso-Oxazole is a colourless liquid, b.p. 96°, and smells like pyridine; it is weakly basic. *iso*-Oxazoles, when substituted in the 3:5-positions, are stable to alkalis, but when the 3-position is vacant, the ring is opened to form ketonitriles (*cf.* oximes, §§2f, 2g. VI).



§4a. **Oxazoles.** Oxazoles may be prepared by the condensation of acid amides with α-halogenoketones, *e.g.*, acetamide and ω-bromoacetophenone form 2-methyl-4-phenyloxazole; the mechanism of the reaction is not certain, but it may occur through the enol forms.



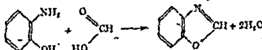
A better method of preparation is the dehydration of α-acylamidocarbonyl compounds with sulphuric acid or phosphorus pentachloride.



Oxazoles have basic properties similar to those of pyridine, but are less resistant to oxidation. They possess aromatic properties, and the stability of the ring towards hydrolytic reagents depends on the nature of the substituents in the ring (*cf. iso*-oxazoles). The parent compound, oxazole, has not yet been prepared.

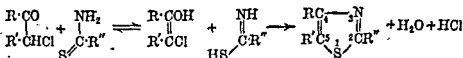
5-Oxazolones. The oxazolones are keto derivatives of the oxazolines, the most important group being the 5-oxazolones or **azlactones.** These azlactones are very important intermediates in the preparation of α-amino-acids (see §2 (*va*), XIII) and keto-acids (see Vol. I).

§4b. **Benzoxazoles.** These may be prepared by the reaction between *o*-aminophenols and carboxylic acids, *e.g.*, *o*-aminophenol and formic acid form benzoxazole, m.p. 31°.

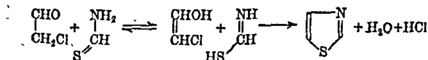


THIAZOLE GROUP

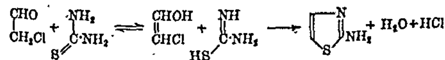
§5. **Thiazoles.** A general method for preparing thiazoles is the condensation between α-halogenocarbonyl compounds (particularly the chloro derivatives) and thioamides; the mechanism of the reaction is uncertain, but it may occur through the enol forms.

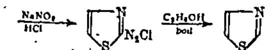


Thiazole itself may be prepared from chloroacetaldehyde and thioformamide.

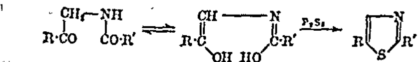


If thiourea or its substitution products are used instead of thioamides, then 2-aminothiazoles are produced, *e.g.*, thiazole may be prepared from chloroacetaldehyde and thiourea as follows:

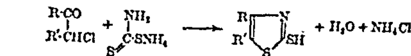




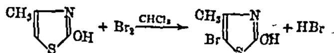
Another general method for preparing thiazoles is by the action of phosphorus pentasulphide on α-acylamidocarbonyl compounds.



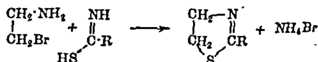
2-Mercaptothiazoles may be prepared by the condensation between α-chloroketones and ammonium dithiocarbamate.



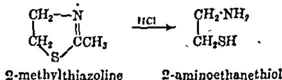
Thiazole is a weakly basic liquid, b.p. 117°; it occurs in vitamin B_1. It is a very stable compound, and is not affected by the usual reducing agents; sodium and ethanol, however, open the ring to form thiols (or hydrogen sulphide) and amines. *Thiazole is very resistant to substitution reactions, but if a hydroxyl group* or an amino group is in position 2, then the molecule is readily attacked by the usual electrophilic reagents to form 5-substitution products, *e.g.*, 2-hydroxy-4-methylthiazole is readily brominated in chloroform solution to *give* 5-bromo-2-hydroxy-4-methyl-*thiazole*.



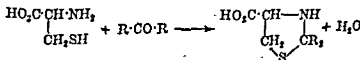
§5a. Thiazolines. These may be prepared by the reaction *between β-halogenoamines and thioamides, e.g.*,



A characteristic reaction of the thiazoles is their ring opening by *the action of acids, e.g.*,

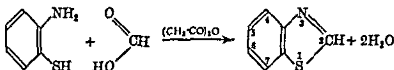


§5b. Thiazolidines. These are readily formed by the condensation of carbonyl compounds with cysteine.

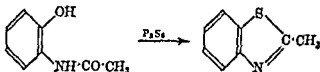


The thiazolidine ring is very easily opened, sometimes by boiling with water, or with an aqueous solution of mercuric chloride (see also penicillin, §6a. XVIII).

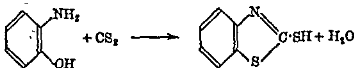
§5c. Benzothiazoles. These may be prepared by the action of acid anhydrides or chlorides on *o*-aminothiophenols, *e.g.*, benzothiazole from *o*-aminothiophenol and formic acid in the presence of acetic anhydride.



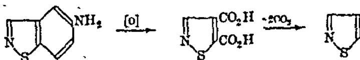
Benzothiazoles are also formed by the action of phosphorus pentasulphide on *o*-acylamidophenols, *e.g.*,



2-Mercaptobenzothiazole is a vulcanisation accelerator (§33a. VIII); it may be prepared as follows:



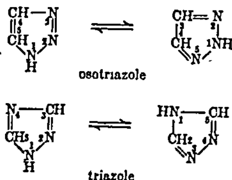
§5d. *iso*Thiazoles. Benz*iso*thiazoles have been known for many years, but no derivatives of *iso*thiazole itself have been obtained until very recently when Adams *et al.* (1956) prepared the parent compound and a number of its simple derivatives, *e.g.*,



*iso*Thiazole is a colourless liquid which smells like pyridine.

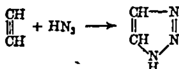
TRIAZOLE GROUP

§6. Osotriazoles and triazoles. Triazoles are five-membered rings which contain two carbon and three nitrogen atoms. Two structural isomeric triazoles are known, the 1:2:3-(1:2:5-) and the 1:2:4- (1:3:4-), the former being known as *osotriazole*, and the latter as *triazole*. Each exists in two dissimilar tautomeric forms.

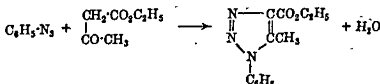


Replacement of the imino hydrogen atom by an alkyl or aryl group prevents tautomerism, and thereby gives rise to the possibility of *two* *N*-substituted triazoles and *two* *N*-substituted osotriazoles. All four types of compounds have been prepared.

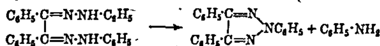
Osotriazole may be prepared by the reaction between acetylene and hydrazoic acid.



On the other hand, a general method for preparing osotriazoles is the condensation of azides with β-ketoesters, *e.g.*, phenyl azide and ethyl acetoacetate form ethyl 5-methyl-1-phenylosotriazole-4-carboxylate.



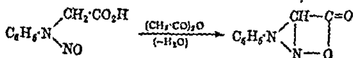
Derivatives of osotriazole may also be prepared by the oxidation of osazones with dichromate and sulphuric acid, or with dilute copper sulphate solution, *e.g.*, benzilosazone gives 1:3:4-triphenylosotriazole.



The formation of osotriazoles from sugar osazones provides a good derivative for the characterisation of sugars (see Vol. I).

Triazoles may be prepared by heating acid hydrazides with amides, *e.g.*, formyl hydrazide and formamide give triazole.

et al. (1935) by the action of cold acetic anhydride on *N*-nitroso-*N*-phenylglycines; Earl formulated the reaction as follows:



Earl (1946) proposed the name *sydnone* for compounds of this type; thus the above compound is *N*-phenylsydnone.

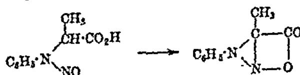
Sydnones are white or pale yellow crystalline compounds, which are hydrolysed by hot 5 per cent. sodium hydroxide to the original *N*-nitroso-*N*-arylglycine, and by moderately concentrated hydrochloric acid to an arylhydrazine, formic acid and carbon dioxide.

The structure proposed by Earl is similar to that of a β-lactone, but Baker *et al.* (1946, 1949) offered a number of objections to this structure, *e.g.*,

(i) A system containing fused three- and four-membered rings would be highly strained, and consequently is unlikely to be produced by dehydration with acetic anhydride; β-lactones are not produced under these conditions.

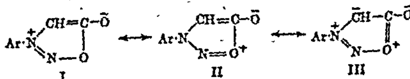
(ii) Many β-lactones are unstable to heat; sydnones are stable and so the β-lactone structure is unlikely.

(iii) If the β-lactone structure is correct, then sydnones should be capable of existing in optically active forms. Kenner and Baker (1946) prepared (+)-*N*-nitroso-*N*-phenylalanine, and when this was converted into a sydnone, the product was optically inactive. If Earl's structure were correct, then the sydnone would be expected to be optically active.

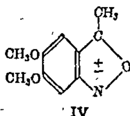


(iv) The aryl nucleus in sydnones is very resistant to substitution by electrophilic reagents. Since the above structure is similar to that of an arylhydrazine, this resistance is unexpected.

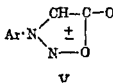
Baker *et al.* (1946) therefore proposed a five-membered ring which cannot be represented by any one purely covalent structure; they put forward a number of charged structures, the sydnone being a resonance hybrid, *e.g.*, three charged resonating structures are:



Now Simpson (1945) had proposed structure IV for 3-methyl-5,6-dimethoxyanthranil; Baker *et al.* (1949) adopted this ± sign and suggested that sydnones be represented by structure V. Baker

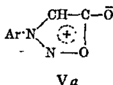


IV



V

also proposed the term **meso-ionic** to describe the sydnone structure. Baker *et al.* (1955) have, however, revised the definition of the term meso-ionic, and have proposed formula V*a* instead of V. This is based on the fact that sydnones are aromatic in

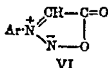


V*a*

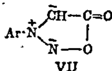
character, and the circle and plus sign represent the sextet of π-electrons in association with a positive charge (the "aromatic sextet" is the essential feature of aromatic compounds).

Dipole moment measurements of various sydnones have shown that the positive end of the dipole is situated on the nitrogen atom attached to the aryl group (Sutton *et al.*, 1947, 1949; Le Fèvre *et al.*, 1947). This is in keeping with Baker's structure.

The meso-ionic structure would necessitate a planar, or almost planar molecule; such a molecule would not be optically active (*cf.* iii above). Earl (1953) has suggested that, from the available evidence, sydnones can be represented as a resonance hybrid, the two main contributing structures being VI and VII.



VI

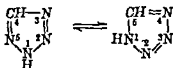


VII

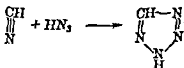
Sutton *et al.* (1949), however, have shown that VI probably contributes to the resonance hybrid, but to a lesser extent than I, II and III.

TETRAZOLE GROUP

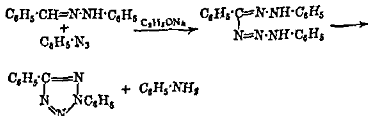
§9. Tetrazole. Tetrazole is a five-membered ring which contains one carbon and four nitrogen atoms. There are two tautomeric forms of tetrazole, and replacement of the imino hydrogen by, *e.g.*, an alkyl group, gives rise to *two* *N*-alkyltetrazoles (*cf.* triazoles, §6).



Tetrazole may be prepared by heating hydrogen cyanide with hydrazoic acid in benzene solution at 110°.



Derivatives of tetrazole may be prepared by the condensation of phenyl azide with phenylhydrazones of aldehydes in the presence of ethanolic sodium ethoxide, *e.g.*, benzaldehyde phenylhydrazone and phenyl azide form 1·4-diphenyltetrazole.



Tetrazole is a colourless solid, m.p. 156°; *it has no basic properties, but the imino hydrogen is acidic, e.g.*, tetrazole forms a silver salt $[CHN_4]^-Ag^+$.

AZINES

The suffix *azine* is used for six-membered rings which contain two or more *hetero-atoms, at least one of which is nitrogen.*

DIAZINE GROUP

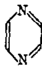
§10. Introduction. The diazines are six-membered rings containing two nitrogen atoms. Three isomeric diazines are theoretically possible, and all three are known.



o-diazine; **pyridazine**

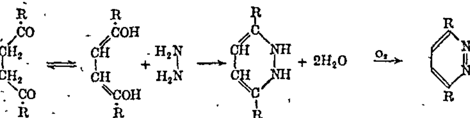


m-diazine; miazine, **pyrimidine**

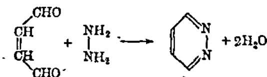


p-diazine, piazine; **pyrazine**

§11. Pyridazines. These may be prepared by the action of hydrazine on 1:4-diketones, the intermediate dihydro compound being readily oxidised by atmospheric oxygen.



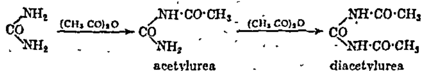
Pyridazine itself may be prepared from maleic dialdehyde and hydrazine hydrate.



Pyridazine is a colourless liquid, b.p. 208°.

PYRIMIDINES

§12. Ureides. Ureides are acylureas, and may be prepared by the action of an acid anhydride or acid chloride on urea, *e.g.*,



The simple ureides resemble the amides in properties.

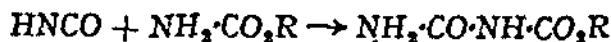
Allophanic acid, $NH_2\cdot CO\cdot NH\cdot CO_2H$, is not known in the free state, but many of its esters have been prepared:

(i) By the action of chloroformates on urea.



$$NH_2\cdot CO\cdot NH_2 + Cl\cdot CO_2R \rightarrow NH_2\cdot CO\cdot NH\cdot CO_2R + HCl$$

(ii) By the reaction between urethans and cyanic acid.



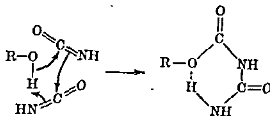
$$HNCO + NH_2\cdot CO_2R \rightarrow NH_2\cdot CO\cdot NH\cdot CO_2R$$

The alkyl allophonates are well-defined crystalline compounds, and so are frequently used to identify alcohols. They are prepared by passing cyanic acid vapour into the dry alcohol; urethans are intermediate products.



$$ROH + HNCO \rightarrow NH_2\cdot CO_2R \xrightarrow{HNCO} NH_2\cdot CO\cdot NH\cdot CO_2R$$

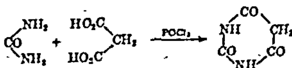
According to Close *et al.* (1953), allophanate formation occurs *via* a concerted attack of two molecules of cyanic acid to form a chelate intermediate.



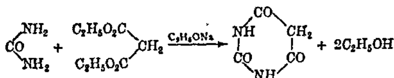
§13. **Cyclic ureides.** Many cyclic ureides are known; some occur naturally and others are synthetic (a number of cyclic ureides —alloxan, allantoin, parabanic acid and hydantoin—are discussed in §2. XVI, *in connection with the purines*, which are cyclic diureides).

The cyclic ureides containing a *six-membered ring* behave, in a number of ways, as *pyrimidine derivatives*.

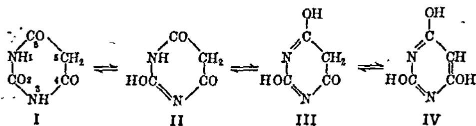
§13a. **Barbituric acid.** A very important pyrimidine derivative is barbituric acid (malonylurea). It was originally prepared by condensing urea with malonic acid in the presence of phosphoryl chloride (Grimaux, 1879).



A much better synthesis is to reflux ethyl malonate with urea in ethanolic solution in the presence of sodium ethoxide.

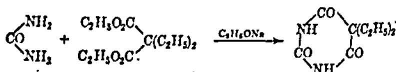


Barbituric acid is a solid, m.p. 253°, and is not very soluble in water. It is strongly acidic due to enolisation (lactam-lactim tautomerism); some possible lactim forms are II–IV. Structure IV represents barbituric acid as 2·4.6-trihydroxypyrimidine, and



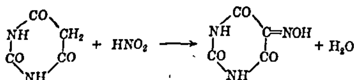
this structure has been proposed because of the acidic nature of barbituric acid. On the other hand, barbituric acid contains an active methylene group, since it readily forms an oximino derivative with nitrous acid. Thus barbituric acid behaves as if it had structure I, II or III. Furthermore, it is very difficult to acylate hydroxypyrimidines containing hydroxyl groups in the 2-, 4- or 6-positions, thus indicating that structure I is more probable than II or III. This is supported by the fact that methylation of hydroxypyrimidines with, *e.g.*, methyl iodide in the presence of sodium hydroxide, results in the formation of *N*-methyl derivatives; this indicates the probable presence of imino groups. On the other hand, it is possible to replace three hydroxyl groups by three chlorine atoms by means of phosphoryl chloride; this suggests barbituric acid behaves as IV. Barbituric acid also forms *O*-alkyl derivatives, thereby indicating structures II, III and IV.

Barbituric acid can be nitrated and brominated in the 5-position, and also forms metallic derivatives (at position 5). By means of the sodio derivative, one or two alkyl groups may be introduced at
the $—CH_2·CO—$ group).
c acid have no hypnotic
arbituric acid (*Barbitone*,
best prepared as follows:



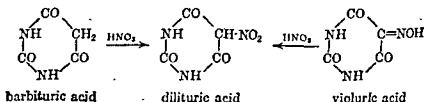
5-*cyclo*Hexyl-3:5-dimethylbarbituric acid (*Evipan*) is a better hypnotic than *Barbitone* and is not so toxic. 5-Ethyl-5-phenylbarbituric acid (*Luminal*) is also used in medicine.

§13b. Derivatives of barbituric acid. **Violuric acid** (5-oximinobarbituric acid) is formed when barbituric acid is treated with nitrous acid; it is the oxime of alloxan (see §2. XVI). Violuric



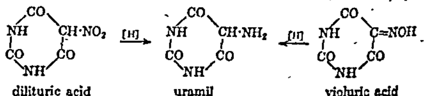
acid gives a violet colour in water, and forms deeply coloured salts with various metals, *e.g.*, the potassium salt is blue and the magnesium and barium salts are purple.

Dilituric acid [illegible] prepared by [illegible] the oxidation [illegible]



barbituric acid dilituric acid violuric acid

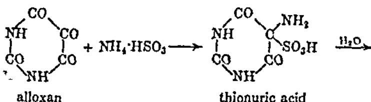
Uramil (5-aminobarbituric acid) is formed by the reduction of either dilituric acid or violuric acid.

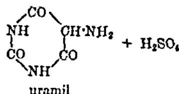


dilituric acid uramil violuric acid

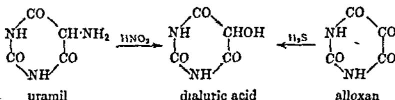
Uramil may also be prepared by the action of ammonium hydrogen

sulphite on alloxan, and then boiling the product, **thionuric acid**, with water.

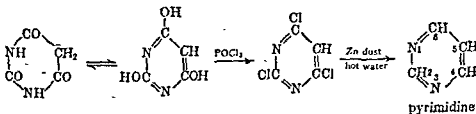




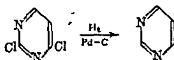
Dialuric acid (5-hydroxybarbituric acid) is produced by the action of nitrous acid on uramil; it is also formed when alloxan is reduced with hydrogen sulphide or with zinc and hydrochloric acid.



§14. **Pyrimidine**, m.p. 22·5°, b.p. 124°/758 mm., was first prepared from barbituric acid as follows (Gabriel, 1900).



Pyrimidine may also be prepared by the oxidation of alkylpyrimidines, followed by decarboxylation. A recent preparation is the catalytic reductive dechlorination of 2:4-dichloropyrimidine; the latter is heated with hydrogen under pressure in the presence of Pd—C and magnesium oxide (Whittaker, 1953).



S

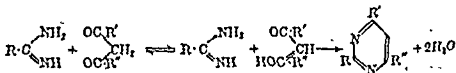
Pyrimidine is neutral in solution, but forms salts with acids. Pyrimidine is probably a resonance hybrid of the following resonating structures:



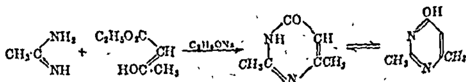
Thus the ring is deactivated, and position 5 has the greatest electron density (*cf.* nitrobenzene and pyridine, Vol. I). It can therefore be expected that attack by electrophilic reagents will be difficult, but attack by nucleophilic reagents (at positions 2, 4 and 6) will be facilitated. Chlorine atoms at 2, 4 or 6 are readily replaced by hydroxyl or amino groups, and an amino group in position 2 or 6 is readily replaced by a hydroxyl group merely on boiling with water (*cf.* vitamin B_1, §3. XVII).

When a hydroxyl or an amino group is present in the pyrimidine nucleus, the compound no longer behaves entirely as an aromatic derivative. The introduction of hydroxyl or amino groups into positions 2, 4 and 6 progressively diminishes the aromatic properties of the compound (*cf.* barbituric acid, §13a, and uracil, §15).

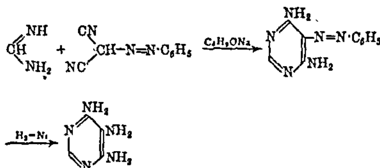
Pyrimidine derivatives. A very important general method for preparing pyrimidines is the condensation between β-carbonyl compounds of the type $R{\cdot}CO{\cdot}CH_2{\cdot}CO{\cdot}R'$, where R and R' = H, R, OR, CN, and compounds having the amidine structure $R{\cdot}C(=NH){\cdot}NH_2$, where R = R (an amidine), OH (urea), SH or SR (thiourea or its *S*-derivative), NH_2 (guanidine); the condensation is carried out in the presence of sodium hydroxide or sodium ethoxide. Thus:



This general reaction may be illustrated by the condensation of acetamidine ($R = CH_3$) with ethyl acetoacetate ($R' = OC_2H_5$, and $R'' = CH_3$) to form 6-hydroxy-2:4-dimethylpyrimidine.

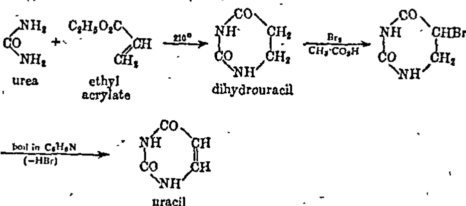


4:5-Diaminopyrimidines, which are intermediates in purine synthesis (see §4. XVI), may be prepared by condensing formamidine with phenylazomalonotrile (Todd *et al.*, 1943).

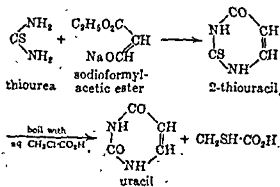


§15. **Uracil** (2:6-dihydroxypyrimidine) is a hydrolytic product of the nucleic acids (§§13, 13b. XVI). It has been synthesised in many ways, *e.g.*,

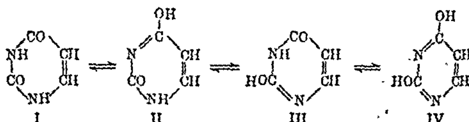
(i) Fischer and Roeder (1901).



(ii) Wheeler and Liddle (1908).



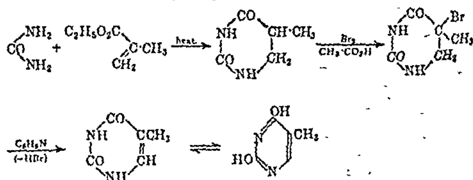
Four tautomeric structures are theoretically possible for uracil.



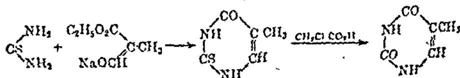
The ultraviolet absorption spectrum of uracil (in ethanol) is different from that of 1:3-dimethyluracil (a derivative of I), from that of 6-methoxy-3-methyluracil (a derivative of II), and from that of 2:6-diethoxyuracil (a derivative of IV). Thus uracil is probably III, and this is supported by the fact that the ultraviolet absorption spectrum of 1-methyluracil (a derivative of III) is similar to that of uracil (Austin, 1934).

§16. **Thymine** (5-methyluracil, 2:6-dihydroxy-5-methylpyrimidine) is a hydrolytic product of the nucleic acids. It has been synthesised by methods similar to those used for uracil.

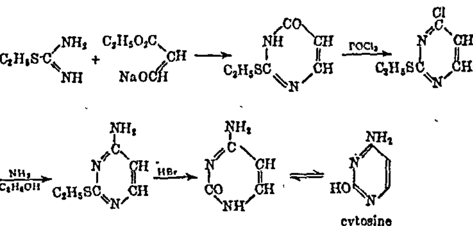
(i) Fischer and Roeder (1901); in this case ethyl methacrylate is used instead of ethyl acrylate.



(ii) Wheeler and Liddle (1908); in this case sodioformylpropionic ester is used instead of sodioformylacetic ester.

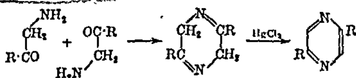


§17. **Cytosine** (6-aminouracil, 6-amino-2-hydroxypyrimidine) is a hydrolytic product of the nucleic acids. It has been synthesised by Wheeler and Johnson (1903) starting from *S*-ethyl*iso*thiourea and sodioformylacetic ester.



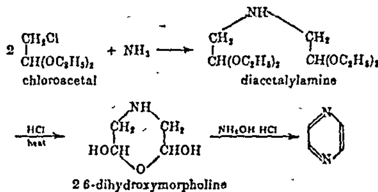
PYRAZINES

§18. [illegible] the self-condensation of an [illegible] oxidising agent such as mer- [illegible] hydro compound is readily [illegible] *al.*, 1893).

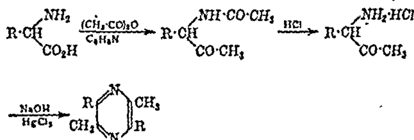


Actually, only the salts of α-aminoketo compounds are known; addition of alkali liberates the free base which immediately forms a pyrazine in the presence of mercuric chloride.

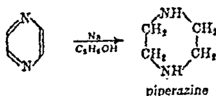
Pyrazine itself may be prepared from aminoacetaldehyde (R = H in the above equations). The best method, however, for preparing pyrazine is as follows (Wolff *et al.*, 1908).



A convenient general method for preparing pyrazines is to heat an α-amino-acid with acetic anhydride in the presence of pyridine, hydrolyse the product (an acetamidoketone) with acid and then warm with sodium hydroxide in the presence of mercuric chloride (Dakin *et al.*, 1928). This method is thus similar to the first general method given above, but offers a convenient method of preparing α-aminocarbonyl compounds.



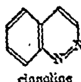
Pyrazine is a solid, m.p. 55°; pyrazines (and pyrazine) are readily reduced by sodium and ethanol to hexahydropyrazines or **piperazines**. Piperazine, m.p. 104°, is a strong diacid base. 2 5-Diketopiperazines are produced from α-amino-acids (see §4 C. XIII).



piperazine

BENZODIAZINES

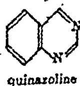
§19. The following benzodiazines are theoretically possible, and all are known; the first two are derived from pyridazine, the third from pyrimidine, and the fourth from pyrazine.



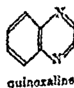
cinnoline



phthalazine

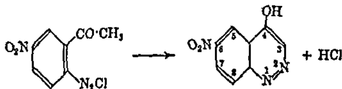


quinazoline

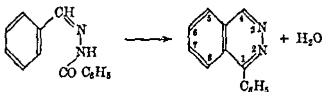


quinoxaline

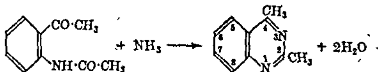
Cinnolines may be prepared by the cyclisation of diazotised *o*-aminoacetophenones (Schofield *et al.*, 1948), *e.g.*,



Phthalazines are formed by heating the benzoyl derivative of benzaldehyde hydrazones, *e.g.*,



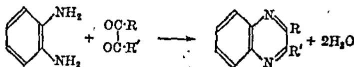
Quinazolines may be prepared by the action of ammonia on acylated *o*-aminobenzaldehydes *or* *o*-aminoacetophenones (Isensee *et al.*, 1948), *e.g.*,



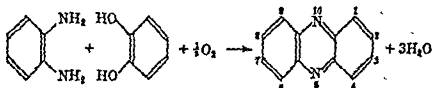
Quinoxalines are formed by the condensation of *o*-phenylenediamines with α-dioxo compounds, *e.g.*,



The formation of quinoxalines is used to identify aromatic *o*-diamines and 1:2-diketones (see, *e.g.*, §9. XVII).



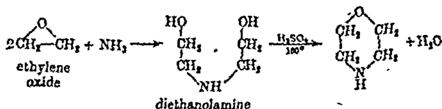
Of the *dibenzodiazines*, only the **phenazines** (*dibenzopyrazines*) are important. Phenazine, m.p. 171°, may be prepared by condensing *o*-phenylenediamine with catechol in the presence of air.



Phenazine forms unstable salts (coloured red or yellow) in excess of strong acids. Many dyes are derived from phenazine, *e.g.*, the safranines (see Vol. I).

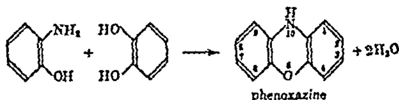
DIAZINES CONTAINING ONE NITROGEN ATOM AND AN OXYGEN OR SULPHUR ATOM

§20. **Oxazines.** *Morpholine* is tetrahydro-1:4-oxazine, and it may be prepared as follows:

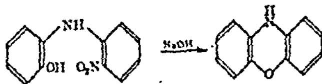


Morpholine is a liquid, b.p. 128°, and is strongly basic. It is miscible with water in all proportions, and is widely used as a solvent.

§21. **Phenoxazines.** These are formed by condensing *o*-aminophenols with catechols at 260°, *e.g.*,

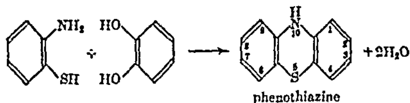


Phenoxazines are also produced by the action of alkali on 2-hydroxy-2'-nitrodiphenylamines, *e.g.*,



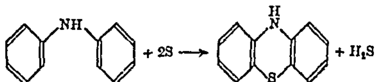
Phenoxazine is a solid, m.p. 156°; it is the parent substance of a number of dyes, *e.g.*, Meldola's Blue (see Vol. I).

§22. Thiazines. *Phenothiazines* may be prepared by heating *o*-aminothiophenols with catechols, *e.g.*,



phenothiazine

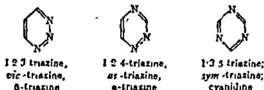
Phenothiazine may also be prepared by fusing diphenylamine with sulphur.



Phenothiazine, m.p. 185°, is used as an insecticide; it is the parent substance of a number of dyes, *e.g.*, Methylene Blue (see Vol. I).

TRIAZINES AND TETRAZINES

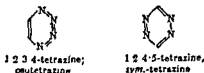
§23. Triazines. Three triazines are theoretically possible; the parent compounds are unknown, but derivatives of each have been prepared.



1 2 3 triazine, *vic*-triazine, β-triazine
1 2 4-triazine, *as*-triazine, α-triazine
1·3 5 triazine; *sym*-triazine; cyanidine

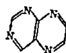
Cyanuric acid, cyamelide and hexamethylenetetramine are derivatives of *sym.*-triazine (see Vol. I).

§24. Tetrazines. Only derivatives of two tetrazines are known.

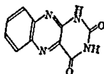


1 2 3 4-tetrazine; osotetrazine
1 2 4·5-tetrazine, *sym.*-tetrazine

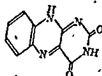
§25. Some important condensed systems containing *two* fused heterocyclic systems are:



pteridine



alloxazine



isoalloxazine

These occur in natural products (see Ch. XVII, Vitamins).

Reading References

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol. IV (1953). Ch. 8 Heterocyclic Chemistry.

Morton, *The Chemistry of Heterocyclic Compounds*, McGraw-Hill (1946).

Patterson and Capell, *The Ring Index*, Reinhold (1940).

Dewar, *The Electronic Theory of Organic Chemistry*, Oxford Press (1949) Pyrazole, pp. 192–193.

Finar and Simmonds, The Reaction between Aroylacetones and Arylhydrazines, *J.C.S.*, **1958**, 200

Wright, The Chemistry of the Benzimidazoles, *Chem. Reviews*, 1951, **48**, 397.

Wiley, The Chemistry of the Oxazoles, *Chem. Reviews*, 1945, **37**, 401.

Organic Reactions, Wiley. Vol. VI (1951). Ch. 8. The Preparation of Thiazoles.

Benson and Savell, The Chemistry of the Vicinal Triazoles, *Chem. Reviews*, 1950, **46**, 1.

Baker and Ollis, Meso-ionic Compounds, *Quart. Reviews (Chem. Soc.)*, 1957, **11**, 15.

Benson, The Chemistry of the Tetrazoles, *Chem. Reviews*, 1947, **41**, 1.

Nineham, The Chemistry of Formazans and Tetrazolium Salts, *Chem. Reviews*, 1955, **55**, 355.

Franklin, Heterocyclic Nitrogen Compounds; Part I. Pentacyclic Compounds, *Chem. Reviews*, 1935, **16**, 305.

Johnson and Hahn, Pyrimidines; Their Amino and Amino-oxy Derivatives, *Chem. Reviews*, 1933, **13**, 193.

Shriner and Neumann, The Chemistry of the Amidines, *Chem. Reviews*, 1944, **35**, p. 395: The formation of substituted pyrimidines.

Lythgoe, Some Aspects of Pyrimidine and Purine Chemistry, *Quart. Reviews (Chem. Soc.)*, 1949, **3**, 181.

Krems and Spoerri, The Pyrazines, *Chem. Reviews*, 1947, **40**, 279.

Leonard, The Chemistry of the Cinnolines, *Chem. Reviews*, 1945, **37**, 269.

Vaughan, The Chemistry of the Phthalazines, *Chem. Reviews*, 1948, **43**, 447.

Gates, The Chemistry of the Pteridines, *Chem. Reviews*, 1947, **41**, 63.

King, Three- and Four-Membered Heterocyclic Rings, *J.C.S.*, 1949, 1318.

CHAPTER XIII

AMINO-ACIDS AND PROTEINS

§1. Classification of the amino-acids. When hydrolysed by acids, alkalis or enzymes, proteins (§6) yield a mixture of amino-acids. In practice it is better to use acid or enzyme hydrolysis th [illegible] e ar [illegible] e ra [illegible] d an improved method for the hydrolysis of proteins. The tissue is first dissolved in 85 per cent. formic acid and then 2N hydrochloric acid is added; *all the amino-acids, except tryptophan, are* liberated within two hours. The nı [illegible] obtained from proteins appears to [illegible]

[illegible]

in any one prevents growth in young animals, and may even cause death. The amino-acids are classified in several ways; the table on pages 545 and 546 shows a convenient classification; the letters *g*, *l* and *e* which follow the name of the acids indicate that the acid is respectively of general occurrence, lesser occurrence, and essential (to man).

The α-amino-acids listed in the table have been isolated from proteins. There are also a few α-amino-acids (and β-) which have been isolated from natural sources, but not from proteins.

§2. General methods of preparation of the amino-acids. There are many general methods for preparing α-amino-acids, but usually each method applies to a small number of particular acids; many acids are also synthesised by methods special to an individual. It should also be noted that very often a synthesis is a more convenient way of preparing an amino-acid than preparing it from natural sources.

(i) **Amination of α-halogenated acids** (Perkin *et al.*, 1858).

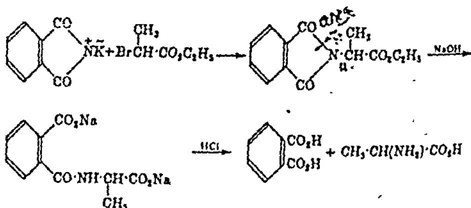
(*a*) An α-chloro- or bromo-acid is treated with concentrated ammonia, *e.g.*,



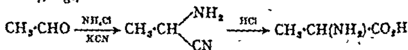
$$CH_2Cl{\cdot}CO_2H + 2NH_3 \rightarrow CH_2(NH_2){\cdot}CO_2H + NH_4Cl$$

This method is convenient for the preparation of glycine, alanine, serine, threonine, valine, leucine and norleucine.

(b) The yields obtained [illegible] riable because [illegible] sing *Gabriel's* [illegible] also Vol. I), [illegible] *e.g.*,

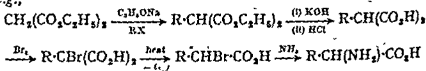


(ii) **Strecker synthesis** (1850). *A cyanohydrin is treated with concentrated ammonia, and the resulting amino-nitrile is then hydrolysed with acid. In practice the amino-nitrile is usually prepared from the oxo compound in one step by treating the latter with an equimolecular mixture of ammonium chloride and potassium cyanide (this mixture is equivalent to ammonium cyanide), e.g.,*



This method is useful for preparing the following amino-acids: glycine, alanine, serine, valine, methionine, glutamic acid, leucine, norleucine and phenylalanine.

(iiia) **Malonic ester synthesis.** This method is really an extension of (i) *a*; it offers a means of preparing α-halogeno-acids, *e.g.*,



This method offers a means of preparing, from readily accessible materials, the following acids: phenylalanine, proline, leucine, isoleucine, norleucine and methionine.

The malonic ester synthesis may also be combined with the

Gabriel phthalimide synthesis to prepare phenylalanine, tyrosine, proline, cystine, serine, aspartic acid, methionine and lysine, *e.g.*,

Cystine.

$$C_6H_5\cdot CH_2SH + HCHO + HCl \longrightarrow C_6H_5CH_2\cdot S\cdot CH_2Cl$$

benzylthiol → benzylthiomethyl chloride

$$CH_2(CO_2C_2H_5)_2 \xrightarrow{Br_2} CHBr(CO_2C_2H_5)_2 \xrightarrow{C_6H_4(CO)_2NK} C_6H_4(CO)_2N\cdot CH(CO_2C_2H_5)_2$$

$$\xrightarrow{Na} C_6H_4(CO)_2N\cdot CNa(CO_2C_2H_5)_2 \xrightarrow{C_6H_5CH_2SCH_2Cl} C_6H_4(CO)_2N\cdot C(CO_2C_2H_5)_2\cdot CH_2\cdot S\cdot CH_2\cdot C_6H_5$$

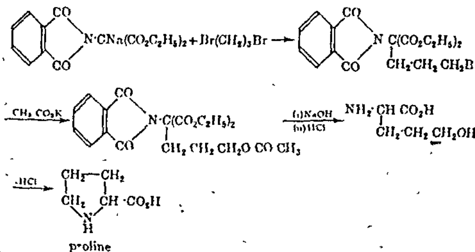
$$\xrightarrow[\text{(ii) HCl}]{\text{(i) NaOH}} NH_2\cdot CH(CO_2H)\cdot CH_2\cdot S\cdot CH_2\cdot C_6H_5 \xrightarrow[\text{liquid } NH_3]{Na} NH_2\cdot CH(CO_2H)\cdot CH_2SH \xrightarrow{\text{air}}$$

S-benzylcysteine; (±)-cysteine

$$NH_2\cdot CH(CO_2H)\cdot CH_2\cdot S\cdot S\cdot CH_2\cdot CH(CO_2H)\cdot NH_2$$

(±)-cystine

Proline.

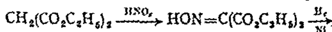


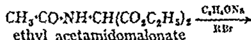
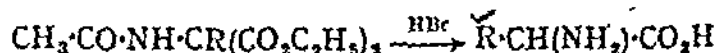
Name	Systematic Name	Formula
Neutral Amino-acids (one amino-group and one carboxyl group)		
1. Glycine (*g*)	Aminoacetic acid	$CH_2(NH_2)\cdot CO_2H$
2. Alanine (*g*)	α-Aminopropionic acid	$CH_3\cdot CH(NH_2)\cdot CO_2H$
3. Valine (*g, e*)	α-Aminoisovaleric acid	$(CH_3)_2CH\cdot CH(NH_2)\cdot CO_2H$
4. Leucine (*g, e*)	α-Amino*iso*caproic acid	$(CH_3)_2CH\cdot CH_2\cdot CH(NH_2)\cdot CO_2H$
5. *iso*Leucine (*g, e*)	α-Amino-β-methyl-*n*-valeric acid	$CH_3\cdot CH_2\cdot CH(CH_3)\cdot CH(NH_2)\cdot CO_2H$
6. Norleucine [1] (*l*)	α-Amino-*n*-caproic acid	$CH_3\cdot(CH_2)_3\cdot CH(NH_2)\cdot CO_2H$
7. Phenylalanine (*g, e*)	α-Amino-β-phenylpropionic acid	$C_6H_5\cdot CH_2\cdot CH(NH_2)\cdot CO_2H$
8. Tyrosine (*g*)	α-Amino-β-(*p*-hydroxyphenyl)propionic acid	$HO\text{-}C_6H_4\text{-}CH_2\cdot CH(NH_2)\cdot CO_2H$
9. Serine (*g*)	α-Amino-β-hydroxypropionic acid	$HOCH_2\cdot CH(NH_2)\cdot CO_2H$
10. Cysteine [2] (*g*)	α-Amino-β-mercaptopropionic acid	$HSCH_2\cdot CH(NH_2)\cdot CO_2H$
11. Cystine (*g*)	Bis-(α-aminopropionic acid)-β-disulphide	$[-SCH_2\cdot CH(NH_2)\cdot CO_2H]_2$
12. Threonine (*g, e*)	α-Amino-β-hydroxy-*n*-butyric acid	$CH_3\cdot CHOH\cdot CH(NH_2)\cdot CO_2H$
13. Methionine (*g, e*)	α-Amino-γ-methylthio-*n*-butyric acid	$CH_3SCH_2\cdot CH_2\cdot CH(NH_2)\cdot CO_2H$
14. Iodogorgic acid (*l*)	3·5-Di-iodotyrosine	$HO\text{-}C_6H_2I_2\text{-}CH_2\cdot CH(NH_2)\cdot CO_2H$
15. Thyroxine (*l*)	β-3 5-Di-iodo-4-(3'.5'-di-iodo-4'-hydroxy)phenyl-α-aminopropionic acid	$HO\text{-}C_6H_2I_2\text{-}O\text{-}C_6H_2I_2\text{-}CH_2\cdot CH(NH_2)\cdot CO_2H$
16. Tryptophan (*g, e*)	α-Amino-β-indolepropionic acid	$C_8H_5NH\text{-}CH_2\cdot CH(NH_2)\cdot CO_2H$ (indole ring)
17. Proline (*g*)	Pyrrolidine-α-carboxylic acid	

17. Proline (*g*)	Pyrrolidine-α-carboxylic acid	CH_2——CH_2 CH_2 $CH{\cdot}CO_2H$ N H
18. Hydroxyproline (*l*)	γ-Hydroxypyrrolidine-α-carboxylic acid	HOCH——CH_2 CH_2 $CH{\cdot}CO_2H$ N H
Acidic Amino-acids (one amino-group and two carboxyl groups)		
19. Aspartic acid (*g*)	α-Aminosuccinic acid	$CO_2H{\cdot}CH_2{\cdot}CH(NH_2){\cdot}CO_2H$
20. Asparagine (*l*)	α-Aminosuccinamic acid	$CONH_2{\cdot}CH_2{\cdot}CH(NH_2){\cdot}CO_2H$
21. Glutamic acid (*g*)	α-Aminoglutaric acid	$HO_2C{\cdot}CH_2{\cdot}CH_2{\cdot}CH(NH_2){\cdot}CO_2H$
22. β-Hydroxyglutamic acid [3]	α-Amino-β-hydroxyglutaric acid	$HO_2C{\cdot}CH_2{\cdot}CHOH{\cdot}CH(NH_2){\cdot}CO_2H$
23. Glutamine (*l*)	α-Aminoglutaramic acid	$CONH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH(NH_2){\cdot}CO_2H$
Basic Amino-acids (two amino-groups and one carboxyl group)		
24. Ornithine [4]	α:δ-Diamino-*n*-valeric acid	$NH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH(NH_2){\cdot}CO_2H$
25 Arginine (*g, e*)	α-Amino-δ-guanidino-*n*-valeric acid	NH_2 $NH{=}C{\cdot}NH{\cdot}(CH_2)_3{\cdot}CH(NH_2){\cdot}CO_2H$
26 Lysine (*g, e*)	α·ε-Diaminocaproic acid	$NH_2{\cdot}(CH_2)_4{\cdot}CH(NH_2){\cdot}CO_2H$
27. Histidine (*g, e*)	α-Amino-β-iminazolepropionic acid	$=CH_2{\cdot}CH(NH_2){\cdot}CO_2H$ HN N

[1] The occurrence of norleucine in proteins is uncertain.
[2] Cysteine has not yet definitely been shown to be present in proteins, but its presence is inferred from various tests.
[3] The occurrence of β-hydroxyglutamic acid in proteins is uncertain.
[4] Ornithine is probably not present in proteins, but is formed by the hydrolysis of arginine.

Acylamido derivatives of malonic ester may also be used to synthesise amino-acids; the usual derivative employed is ethyl acetamidomalonate (Albertson, 1946).

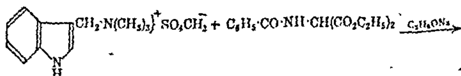


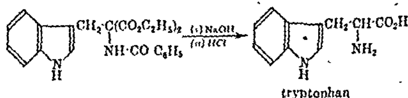



$$CH_2(CO_2C_2H_5)_2 \xrightarrow{HNO_2} HON{=}C(CO_2C_2H_5)_2 \xrightarrow[Ni]{H_2}$$
$$NH_2{\cdot}CH(CO_2C_2H_5)_2 \xrightarrow{CH_3COCl}$$
$$CH_3{\cdot}CO{\cdot}NH{\cdot}CH(CO_2C_2H_5)_2 \xrightarrow[RBr]{C_2H_5ONa}$$
ethyl acetamidomalonate
$$CH_3{\cdot}CO{\cdot}NH{\cdot}CR(CO_2C_2H_5)_2 \xrightarrow{HBr} R{\cdot}CH(NH_2){\cdot}CO_2H$$

The following acids may be prepared by this method: serine, leucine, valine, methionine, lysine, glutamic acid and ornithine.

A special application of this method is the preparation of tryptophan from benzamidomalonic ester and gramine methosulphate (Albertson *et al.*, 1945; Tishler *et al.*, 1945).

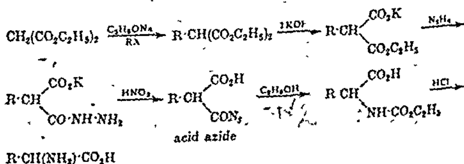




tryptophan

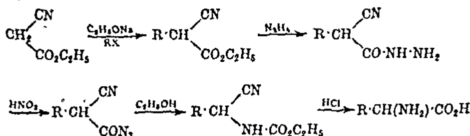
A more recent method of preparing ethyl acetamidomalonate is to reduce oximinomalonic ester in a mixture of acetic anhydride, pyridine and sodium acetate with hydrogen in the presence of Raney nickel (Vignau, 1952).

(*iiib*). *α-Amino-acids may be synthesised* by means of the **Curtius reaction** (see also Vol. I).



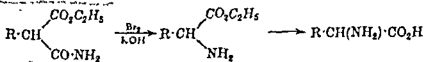
Glycine, alanine, phenylalanine and valine can be prepared by this method.

Instead of malonic ester, the starting material can be ethyl *cyanoacetate*.

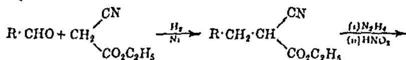


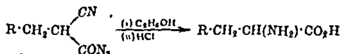
Phenylalanine and tyrosine are conveniently prepared by this method.

Another variation is the use of the Hofmann degradation on ester amides (see also Vol. I).

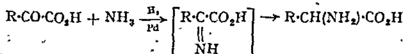


(iiic) **The Darapsky synthesis** (1936). In this method an *aldehyde is condensed with ethyl cyanoacetate and simultaneously* hydrogenated; the product, an alkylcyanoacetic ester, is then treated as above (for the cyanoacetic ester method).



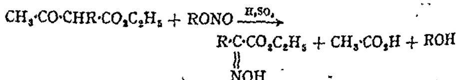


(iv) Amino-acids may be prepared by reducing α-ketonic acids in the presence of ammonia; the reduction may be performed catalytically, or with sodium and ethanol. The mechanism of the reaction is not certain, but it probably occurs *via* the imino-acid.



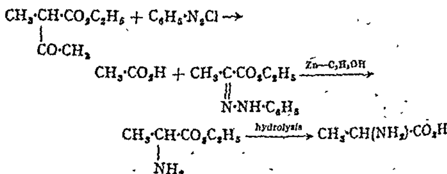
This method works well for alanine and glutamic acid.

Oximes of α-keto-acids may also be reduced to α-amino-acids. The advantage of this method is that the oximes may readily be prepared in good yield by the action of sulphuric acid on a mixture of an alkylacetoacetic ester and an alkyl nitrite (Hartung *et al.*, 1942).



$$CH_3{\cdot}CO{\cdot}CHR{\cdot}CO_2C_2H_5 + RONO \xrightarrow{H_2SO_4} R{\cdot}C(=NOH){\cdot}CO_2C_2H_5 + CH_3{\cdot}CO_2H + ROH$$

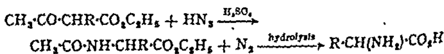
The reduction of phenylhydrazones made by the action of a diazonium salt on an alkylacetoacetic ester also may be used to prepare α-amino-acids (*cf.* the Japp–Klingermann reaction, Vol. I); *e.g.*,



$$CH_3{\cdot}CH(CO{\cdot}CH_3){\cdot}CO_2C_2H_5 + C_6H_5{\cdot}N_2Cl \rightarrow CH_3{\cdot}CO_2H + CH_3{\cdot}C(=N{\cdot}NH{\cdot}C_6H_5){\cdot}CO_2C_2H_5 \xrightarrow{Zn-C_2H_5OH}$$

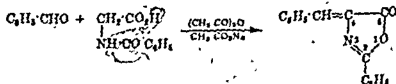
$$CH_3{\cdot}CH(NH_2){\cdot}CO_2C_2H_5 \xrightarrow{hydrolysis} CH_3{\cdot}CH(NH_2){\cdot}CO_2H$$

[illegible] *ne and hydroxy-* [illegible] to α-amino-acids by means of the Schmidt reaction (see also Vol. I).



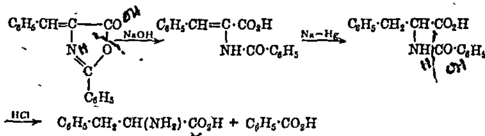
$$CH_3{\cdot}CO{\cdot}CHR{\cdot}CO_2C_2H_5 + HN_3 \xrightarrow{H_2SO_4} CH_3{\cdot}CO{\cdot}NH{\cdot}CHR{\cdot}CO_2C_2H_5 + N_2 \xrightarrow{hydrolysis} R{\cdot}CH(NH_2){\cdot}CO_2H$$

(va). **The Azlactone synthesis** (*Erlenmeyer synthesis*, 1893). Azlactones are usually prepared by heating an aromatic aldehyde with hippuric acid (benzoylglycine) in the presence of acetic anhydride and sodium acetate, *e.g.*, benzaldehyde forms benzoyl-α-aminocinnamic azlactone (4-benzylidene-2-phenyloxazol-5-one).



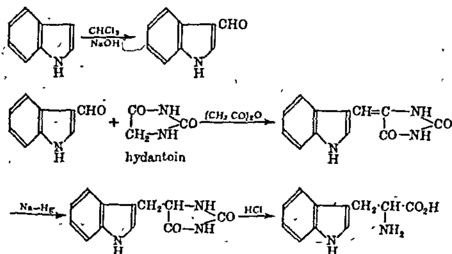
This reaction is usually referred to as the **Erlenmeyer azlactone synthesis.** Aceturic acid (acetylglycine) may also be used instead of hippuric acid. Furthermore, it has been found that aliphatic aldehydes may condense with hippuric acid to form azlactones if lead acetate is used instead of sodium acetate (Finar *et al.*, 1949).

When azlactones are warmed with one per cent. sodium hydroxide solution, the ring is opened, and if the product is reduced with sodium amalgam followed by hydrolysis with acid, an α-amino-acid is produced, *e.g.*,

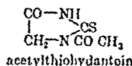


The azlactone synthesis offers a convenient means of preparing phenylalanine, tyrosine, tryptophan and thyroxine.

(*vb*) Aromatic aldehydes also condense with hydantoin, and reduction of the product with sodium amalgam or ammonium hydrogen sulphide, followed by hydrolysis, gives an α-amino-acid, *e.g.*, tryptophan may be prepared by first converting indole into indole-3-aldehyde by means of the Reimer–Tiemann reaction (see Vol. I).

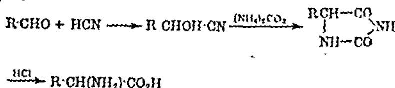


This method has been improved by using acetylthiohydantoin instead of hydantoin.

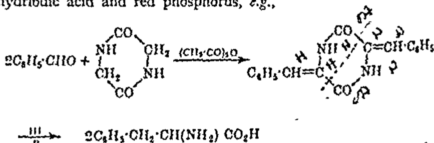


The above method may be used to prepare phenylalanine, tyrosine, tryptophan and methionine.

Another modification of the hydantoin synthesis is the **Bücherer hydantoin synthesis** (1934). In this method an oxo compound is converted into the cyanohydrin and this, on treatment with ammonium carbonate, produces a 5-substituted hydantoin which, on hydrolysis, gives an α-amino-acid.



(vc) Aromatic aldehydes may be condensed with diketopiperazine, and the product converted into an amino-acid by heating with hydriodic acid and red phosphorus, *e.g.*,



Phenylalanine, tyrosine and methionine may be prepared by this method.

§3. Isolation of amino-acids from protein hydrolysates. Many amino-acids can be detected colorimetrically, and these colour reactions have now been developed for quantitative estimation. Also, amino-acids containing a benzene or pyrrolidine nucleus have characteristic absorption spectra; thus the presence of such acids can readily be ascertained.

The actual *quantitative isolation* of amino-acids from their mixtures is a difficult problem. The earliest method was the fractional distillation of the amino-acid esters *in vacuo* (Fischer, 1901). This method is very little used now, and is only satisfactory for the

neutral amino-acids (*i.e.*, those containing one amino-group and one carboxyl group).

Neutral amino-acids may be extracted by *n*-butanol saturated with water, and then separated by fractional crystallisation or by the fractional distillation of the esters. After the butanol extraction, the residue may be treated with phosphotungstic acid, whereupon the *basic* amino-acids are precipitated (Dakin *et al.*, 1913).

A number of individual amino-acids can be obtained by means of selective precipitation as salts, *e.g.*, lysine is precipitated by picric acid.

Mixtures of amino-acids may be separated into fractions consisting of the neutral, basic and acidic acids by means of the *electrical transport method*. In this method a P.D. is applied to the mixture at the proper pH; the basic acids (positively charged) migrate to the cathode compartment, the acidic acids (negatively charged) migrate to the anode compartment, and the neutral acids remain in the centre compartment.

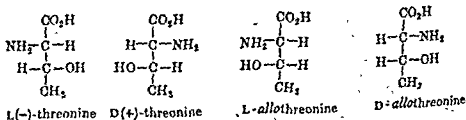
The most satisfactory general method of separating the amino-acids is by means of partition chromatography of the acetamido-acids (Martin *et al.*, 1941). The dinitrodiphenylamino-acids have also been separated in this way (Sanger *et al.*, 1945, 1948; see also §11).

A very interesting analytical method is the *microbiological assay*. This depends on the fact that micro-organisms can be "trained" to feed on a specific amino-acid in the nutrient medium. The rate of growth of the micro-organism is first measured by breeding in a medium containing the particular amino-acid, and then the rate of growth is measured in the mixture of amino-acids to be analysed. In this way it is possible to determine the amounts of various amino-acids in protein hydrolysates *without* isolation of the acids. Another method of analysis is that of *isotopic dilution*. Suppose the amount of glycine is to be estimated. A weighed amount of labelled glycine is added to the hydrolysate, and then glycine is isolated by one of the standard methods. The amount of *labelled* glycine in this specimen is now measured, *e.g.*, say it is 1 per cent. Thus for every 1 g. of labelled glycine there are 99 g. of ordinary glycine. Since the weight of the added labelled glycine is known, the total weight of glycine in the mixture can therefore be calculated (see also Vol. I).

§4. General properties of the amino-acids. The amino-acids are colourless crystalline compounds which are generally soluble in water but sparingly soluble in organic solvents; most melt with decomposition, but Gross *et al.* (1955) have shown that

sublimation is possible with a number of amino-acids. All except glycine contain at least one asymmetric carbon atom, and all (except glycine) occur naturally in their optically active forms. It has been shown that the α-carbon atom, *i.e.*, the carbon atom attached to the amino-group, has, in almost all the amino-acids, the same configuration as L(−)-glyceraldehyde. The specific rotation of the amino-acids depends on the *p*H of the solution, the temperature, the presence of salts and the nature of the solvent (see §12. I). The racemic amino-acids may be resolved by first formylating and then resolving the formyl derivatives *via* the salt with an optically active base, and finally removing the formyl group by hydrolysis (see also C i). Alternatively, racemic amino-acids may be resolved by means of enzymes (see §10 iv. II). As pointed out above, the natural amino-acids are L; these are obtained by acid or enzymic hydrolysis of proteins. Alkaline hydrolysis of proteins gives the DL-amino-acids (§1), and so does the synthetic preparation; it is by resolution of the synthetic racemic modification that the D-amino-acids are frequently prepared.

The symbols D and L are used for the configuration of the α-carbon atom (see above), and the symbols (+) and (−) are used to indicate the direction of the rotation (*cf.* §5. II). When two asymmetric centres are present, then D and L still refer to the α-carbon atom, and the *naturally occurring acid* is known as the L-amino-acid. The *allo*-form is the name given to that form in which the configuration of the *second asymmetric carbon atom is inverted*, *e.g.*, L(−)-threonine (the naturally occurring form), D(+)-threonine, L-*allo*threonine and D-*allo*threonine.



L(−)-threonine D(+)-threonine L-*allo*threonine D-*allo*threonine

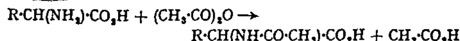
Since they contain amino and carboxyl groups, the amino-acids possess the properties of both a base and an acid, *i.e.*, they are amphoteric.

A. Reactions due to the amino-group.

(i) The amino-acids form salts with strong inorganic acids, *e.g.*, $\overset{-}{Cl}\{H_3\overset{+}{N}\cdot CH_2\cdot CO_2H$. These salts are usually sparingly soluble in

water, and the free acid may be liberated from its salt by means of a strong organic base, *e.g.*, pyridine.

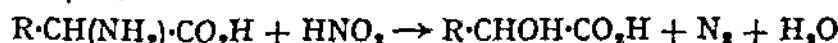
(ii) Amino-acids may be acetylated by means of acetyl chloride or acetic anhydride.



$$R{\cdot}CH(NH_2){\cdot}CO_2H + (CH_3{\cdot}CO)_2O \rightarrow R{\cdot}CH(NH{\cdot}CO{\cdot}CH_3){\cdot}CO_2H + CH_3{\cdot}CO_2H$$

Similarly, benzoyl chloride produces the benzoyl derivative. These acylated derivatives are acidic, the basic character of the amino-group being effectively eliminated by the presence of the negative group attached to the nitrogen. It should also be noted that the carboxyl group of one molecule can react with the amino-group of another molecule of an amino-acid to form a peptide (see §9). Sanger (1945) has shown that 1-fluoro-2:4-dinitrobenzene combines with amino-acids to form dinitrophenyl derivatives (see §11).

(iii) Nitrous acid liberates nitrogen from amino-acids.



$$R{\cdot}CH(NH_2){\cdot}CO_2H + HNO_2 \rightarrow R{\cdot}CHOH{\cdot}CO_2H + N_2 + H_2O$$

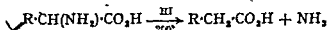
The nitrogen is evolved quantitatively, and this forms the basis of the **van Slyke method** (1911) for analysing mixtures of amino-acids.

(iv) Nitrosyl chloride (or bromide) reacts with amino-acids to form chloro- (or bromo) acids.



$$R{\cdot}CH(NH_2){\cdot}CO_2H + NOCl \rightarrow R{\cdot}CHCl{\cdot}CO_2H + N_2 + H_2O$$

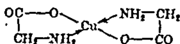
(v) When heated with hydriodic acid at 200°, the amino-group is eliminated with the formation of a fatty acid.



$$R{\cdot}CH(NH_2){\cdot}CO_2H \xrightarrow[200^\circ]{HI} R{\cdot}CH_2{\cdot}CO_2H + NH_3$$

B. Reactions due to the carboxyl group.

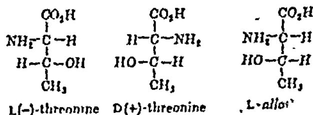
(i) Amino-acids form salts; the salts of the heavy metals are chelate compounds, *e.g.*, the copper salt of glycine (deep blue needles) is formed by heating copper oxide with an aqueous solution of glycine.



The amino-acids may be liberated from their alkali salts by treatment in ethanolic solution with ethyl oximinocyanoacetate (Galat, 1947).

sublimation is possible with a number of amino-acids. All exc glycine contain at least one asymmetric carbon atom, and all (cept glycine) occur naturally in their optically active forms. It been shown that the α-carbon atom, *i.e.*, the carbon atom attac to the amino-group, has, in almost all the amino-acids, the s configuration as L(−)-glyceraldehyde. The specific rotation of amino-acids depends on the pH of the solution, the temperat the presence of salts and the nature of the solvent (see § The racemic amino-acids may be resolved by first formy then resolving the formyl derivatives *via* the salt with an active base, and finally removing the formyl group by (see also C i). Alternatively, racemic amino-acids may be by means of enzymes (see §10 iv. II). As pointed out natural amino-acids are L; these are obtained by acid or hydrolysis of proteins. Alkaline hydrolysis of proteins DL-amino-acids (§1), and so does the synthetic preparati by resolution of the synthetic racemic modification D-amino-acids are frequently prepared.

The symbols D and L are used for the configuration of atom (see above), and the symbols (+) and (−) are used the direction of the rotation (*cf.* §5. II). When two centres are present, then D and L still refer to the and the *naturally occurring acid* is known as the L *allo*-form is the name given to that form in which the of the *second asymmetric carbon atom is inverted, e.g.*, (the naturally occurring form), D(+)-threonine, L-*all* D-*allo*threonine.



L(−)-threonine D(+)-threonine L-allo

Since they contain amino and carboxyl possess the properties of both a base and amphoteric.

A. Reactions due to the amino

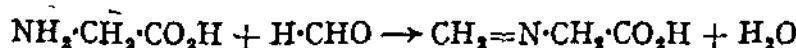
(i) The amino-acids form salts with

$\bar{Cl}\{H_3\overset{+}{N}\cdot CH_2\cdot CO_2H$. These salts are

melting point and its insolubility in hydrocarbon solvents, also indicate that it exists as the inner salt in the solid state.

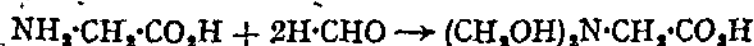
Each amino-acid has a definite pH at which it does not migrate to either electrode when a P.D. is applied. This pH is known as the **isoelectric point**, and at this point the amino-acid has its lowest solubility.

Owing to their amphoteric character, amino-acids cannot be titrated directly with alkali. When formalin solution is added to glycine, methyleneglycine is formed.



$$NH_2{\cdot}CH_2{\cdot}CO_2H + H{\cdot}CHO \longrightarrow CH_2{=}N{\cdot}CH_2{\cdot}CO_2H + H_2O$$

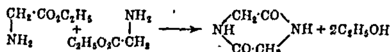
Although some methyleneglycine is probably formed, it appears that the reaction is more complex; the main product appears to be dimethylolglycine.



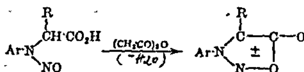
$$NH_2{\cdot}CH_2{\cdot}CO_2H + 2H{\cdot}CHO \longrightarrow (CH_2OH)_2N{\cdot}CH_2{\cdot}CO_2H$$

These glycine derivatives are strong acids (the basic character of the amino-group being now suppressed), and can be titrated with alkali. This method is known as the *Sörensen formol titration.*

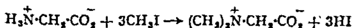
(ii) When heated, α-amino-acids form 2:5-diketopiperazines; esters give better yields; *e.g.*, diketopiperazine from glycine ester.



(iii) N-alkyl or arylamino-acids form N-nitroso derivatives with nitrous acid, and these may be dehydrated to sydnones by means of acetic anhydride (see §8. XII).

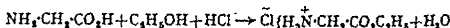


(iv) **Betaines.** These are the trialkyl derivatives of the amino-acids; betaine itself may be prepared by heating glycine with methyl iodide in methanolic solution. The betaines exist as dipolar ions; thus the formation of betaine may be written:



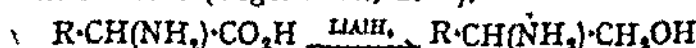
$$H_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2^- + 3CH_3I \longrightarrow (CH_3)_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2^- + 3HI$$

(ii) When heated with an alcohol in the presence of dry hydrogen chloride, amino-acids form ester hydrochlorides, *e.g.*,



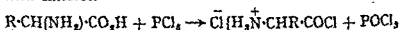
$$NH_2 \cdot CH_2 \cdot CO_2H + C_2H_5OH + HCl \longrightarrow \bar{Cl}\{H_3\overset{+}{N} \cdot CH_2 \cdot CO_2C_2H_5 + H_2O$$

The free ester may be obtained by the action of aqueous sodium carbonate on the ester salt. The esters are fairly readily hydrolysed to the amino-acid by aqueous sodium hydroxide (even at room temperature). These esters may be reduced to the amino-alcohols by means of sodium and ethanol, or hydrogenated in the presence of Raney nickel. Amino-acids may be reduced directly to the amino-alcohol with lithium aluminium hydride, and in this case no racemisation occurs (Vogel *et al.*, 1952).



$$R \cdot CH(NH_2) \cdot CO_2H \xrightarrow{LiAlH_4} R \cdot CH(NH_2) \cdot CH_2OH$$

(iii) When suspended in acetyl chloride and then treated with phosphorus pentachloride, amino-acids form the hydrochloride of the acid chloride.



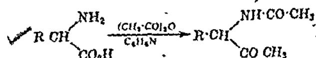
$$R \cdot CH(NH_2) \cdot CO_2H + PCl_5 \longrightarrow \bar{Cl}\{H_3\overset{+}{N} \cdot CHR \cdot COCl + POCl_3$$

(iv) Dry distillation, or better by heating with barium oxide, decarboxylates amino-acids to amines.



$$R \cdot CH(NH_2) \cdot CO_2H \longrightarrow R \cdot CH_2 \cdot NH_2 + CO_2$$

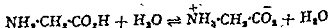
(v) When heated with acetic anhydride in pyridine solution, amino-acids are converted into methyl α-acetamidoketones (Dakin *et al.*, 1928; see also §18. XII).



$$R \cdot CH(NH_2)(CO_2H) \xrightarrow[C_5H_5N]{(CH_3 \cdot CO)_2O} R \cdot CH(NH \cdot CO \cdot CH_3)(CO \cdot CH_3)$$

C. Reactions due to both the amino and carboxyl groups.

(i) When measured in aqueous solution, the dipole moment of glycine (and other amino-acids) is found to have a large value. To account for this large value it has been suggested that glycine exists, in solution, as an *inner salt*:



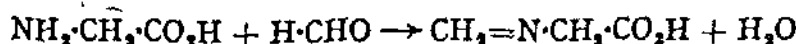
$$NH_2 \cdot CH_2 \cdot CO_2H + H_2O \rightleftharpoons \overset{+}{N}H_3 \cdot CH_2 \cdot C\bar{O}_2 + H_2O$$

Such a doubly charged ion is also known as a *zwitterion*, *ampholyte*, or a *dipolar ion*. This dipolar ion structure also accounts for the ... (the carboxyl ... h other to form a salt). The properties ... its high

melting point and its insolubility in hydrocarbon solvents, also indicate that it exists as the inner salt in the solid state.

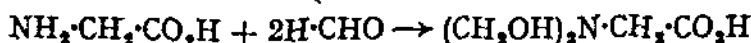
Each amino-acid has a definite *p*H at which it does not migrate to either electrode when a P.D. is applied. This *p*H is known as the **isoelectric point**, and at this point the amino-acid has its lowest solubility.

Owing to their amphoteric character, amino-acids cannot be titrated directly with alkali. When formalin solution is added to glycine, methyleneglycine is formed.



$$NH_2{\cdot}CH_2{\cdot}CO_2H + H{\cdot}CHO \longrightarrow CH_2{=}N{\cdot}CH_2{\cdot}CO_2H + H_2O$$

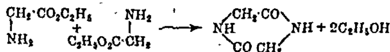
Although some methyleneglycine is probably formed, it appears that the reaction is more complex; the main product appears to be dimethylolglycine.



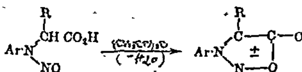
$$NH_2{\cdot}CH_2{\cdot}CO_2H + 2H{\cdot}CHO \longrightarrow (CH_2OH)_2N{\cdot}CH_2{\cdot}CO_2H$$

These glycine derivatives are strong acids (the basic character of the amino-group being now suppressed), and can be titrated with alkali. This method is known as the *Sörensen formol titration.*

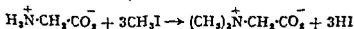
(ii) When heated, α-amino-acids form 2:5-diketopiperazines; esters give better yields; *e.g.*, diketopiperazine from glycine ester.



(iii) *N*-alkyl or arylamino-acids form *N*-nitroso derivatives with nitrous acid, and these may be dehydrated to sydnones by means of acetic anhydride (see §8. XII).

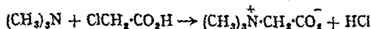


(iv) **Betaines.** These are the trialkyl derivatives of the amino-acids; betaine itself may be prepared by heating glycine with methyl iodide in methanolic solution. The betaines exist as dipolar ions; thus the formation of betaine may be written:



$$H_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2^- + 3CH_3I \longrightarrow (CH_3)_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2^- + 3HI$$

Betaine is more conveniently prepared by warming an aqueous solution of chloroacetic acid with trimethylamine.

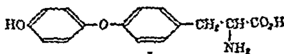


$$(CH_3)_3N + ClCH_2{\cdot}CO_2H \rightarrow (CH_3)_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2^- + HCl$$

Betaine is a solid, m.p. 300° (with decomposition). It occurs in nature, especially in plant juices. It behaves as a base, *e.g.*, with hydrochloric acid it forms the stable crystalline hydrochloride, $\overset{-}{Cl}\{(CH_3)_3\overset{+}{N}{\cdot}CH_2{\cdot}CO_2H$.

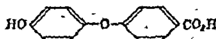
§5. Thyroxine (thyroxin). Thyroxine is a *hormone*; it is the active principle of the thyroid gland and was first isolated by Kendall (1919). It was later isolated by Harington (1930) as a white crystalline solid, m.p. 235°, with a lævorotation.

The structure of thyroxine was established by Harington (1926). This author showed that the molecular formula of thyroxine is $C_{15}H_{11}O_4NI_4$. When treated in alkaline solution with hydrogen in the presence of colloidal palladium, the iodine in thyroxine is replaced by hydrogen to form thyronine (thyronin), $C_{15}H_{15}O_4N$. This behaves as a phenol and an α-amino-acid. On fusion with potassium hydroxide in an atmosphere of hydrogen, thyronine gives a mixture of *p*-hydroxybenzoic acid, quinol, oxalic acid and ammonia. When fused with potassium hydroxide at 250°, thyronine gives *p*-hydroxybenzoic acid, quinol and a compound with the molecular formula $C_{13}H_{12}O_2$ (II). A structure for thyronine which would give all these products is I.



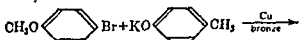
thyronine

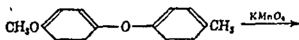
Thyronine (provisionally structure I) was subjected to the Hofmann exhaustive methylation (see §4. XIV) and the product thereby obtained was then oxidised. The final product would be III (on the assumption that I is thyronine).

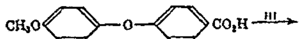


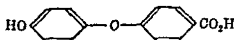
III

The structure of III was confirmed by synthesis, starting from *p*-bromoanisole and *p*-cresol.



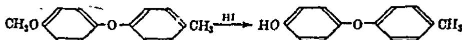






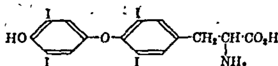
III

Furthermore, when 4-methoxy-4′-methyldiphenyl ether is heated with hydriodic acid, compound II ($C_{13}H_{12}O_2$; see above) is obtained; thus the structure of II is also established.



II

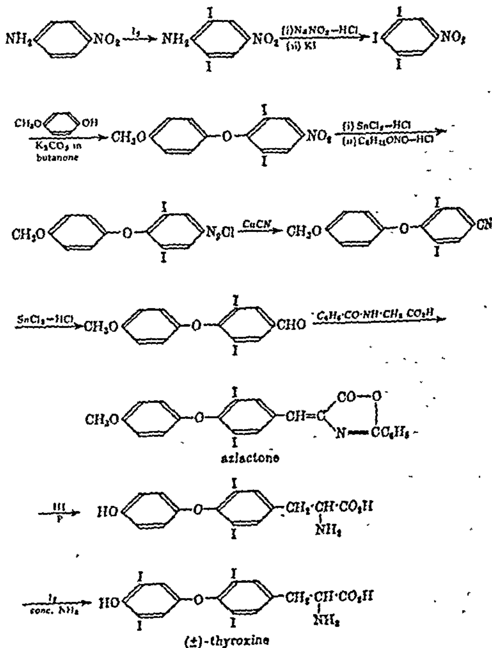
Now when thyroxine is fused with potassium hydroxide, no *p*-hydroxybenzoic acid is obtained; instead, compounds of the pyrogallol type are formed. These facts suggest that two atoms of iodine are adjacent to the hydroxyl group, and that the two remaining iodine atoms are in the other benzene ring. This, together with the analogy with di-iodotyrosine, leads to the suggestion that thyroxine is IV.



IV

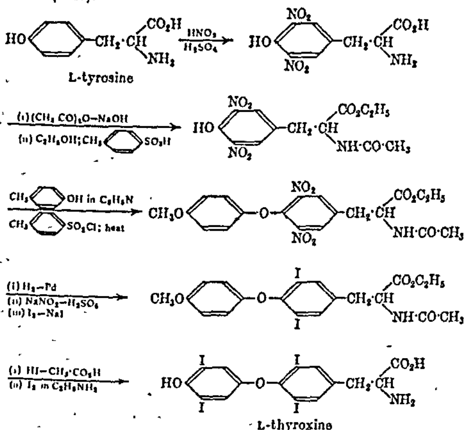
thyroxine

This structure for thyroxine has been confirmed by synthesis (Harington *et al.*, 1927).



The racemic modification was resolved *via* the formyl derivative (Harington, 1938).

The synthesis of thyroxine has been improved, *e.g.*, by Hems *et al.* (1949).



The thyroid gland also contains 3·5-di-iodotyrosine, and this compound is believed to be the precursor of thyroxine.

PROTEINS

§6. General nature of proteins. The name *protein* was introduced by Mulder (1839), who derived it from the Greek word *proteios* (meaning *first*).
occur in the protoplasm (
position varies with the s
be given as: carbon, 47–50%; hydrogen, 6–7%; oxygen, 24–25%; nitrogen, 16–17%; sulphur, 0·2–0·3%. Other elements may also be present, *e.g.*, phosphorus (nucleoproteins), iron (hæmoglobin).

Proteins are colloids and have no characteristic melting points; some have been obtained in crystalline form. All proteins are optically active (lævorotatory), their activity arising from the fact

that they are complex substances built up of amino-acids. It appears likely that all enzymes are proteins (see §12); many hormones are also proteins, *e.g.*, insulin.

Proteins may be coagulated, *i.e.*, precipitated irreversibly, by heat and by strong inorganic acids and bases, etc. When proteins are precipitated irreversibly, they are said to be *denatured*, but the chemical changes that occur in this process are still uncertain. Proteins may be precipitated by ethanol or concentrated solutions of ammonium sulphate or sodium chloride. In this case, the precipitation is reversible, *i.e.*, the precipitated proteins may be redissolved; thus they are not denatured by these reagents. Proteins are also precipitated by the salts of the heavy metals, *e.g.*, mercuric chloride, copper sulphate, etc., and they give many characteristic colour reactions with various reagents, *e.g.*,

(i) *Biuret reaction.* Addition of a very dilute solution of copper sulphate to an alkaline solution of a protein produces a red or violet colour.

(ii) *Millon's reaction.* When a solution of mercuric nitrate containing nitrous acid is added to a protein solution, a white precipitate is formed and slowly turns pink.

(iii) *Xanthoproteic reaction.* Proteins produce a yellow colour when treated with concentrated nitric acid.

Proteins are amphoteric, their behaviour as an anion or a cation depending on the *p*H of the solution. At some definite *p*H, characteristic for each protein, the solution contains equal amounts of anion and cation. In this condition the protein is said to be at its *isoelectric point*, and at this *p*H the protein has its least solubility, *i.e.*, it is most readily precipitated (*cf.* amino-acids, §4. C i). The osmotic pressure and viscosity of the protein solution are also a minimum at the isoelectric point. The amphoteric nature of proteins is due to the presence of a large number of free acidic and basic groups arising from the amino-acid units in the molecule. These groups can be titrated with alkali or acid, and by this means it has been possible to identify acidic and basic groups belonging to the various amino-acid units (see also §11).

The molecular weights of proteins have been determined by means of the ultracentrifuge, osmotic pressure measurements, X-ray diffraction, light scattering effects, and by chemical analysis. Chemical methods are based on the estimation of a particular amino-acid, *e.g.*, casein contains cystine; hence the estimation of the percentage of this amino-acid and of sulphur will lead to the evaluation of the molecular weight of casein. The most reliable values of the molecular weights are those obtained by the ultracentrifuge method; the values recorded vary considerably for the

individual proteins, ranging from about 40,000 for egg albumin to about 5,000,000 for hæmocyanin.

§7. Classification of proteins. Several arbitrary classifications of the proteins are in use. One method is based mainly on physical properties, particularly solubility.

A. Simple proteins. These give only amino-acids on hydrolysis.

(i) *Albumins.* These are soluble in water (and in acids and alkalis), and are coagulated by heat. They are precipitated by saturating their solutions with ammonium sulphate.

Albumins are usually low or deficient in glycine; some albumins are serum albumin, egg albumin and lactalbumin.

(ii) *Globulins.* These are insoluble in water, but are soluble in dilute salt solution and in dilute solutions of strong inorganic acids and alkalis. They are precipitated by half saturating their solutions with ammonium sulphate, and they are coagulated by heat.

Globulins usually contain glycine; some typical globulins are serum globulin, tissue globulin and vegetable globulin.

(iii) *Prolamines.* These are insoluble in water or salt solution, but are soluble in dilute acids and alkalis, and in 70–90 per cent. ethanol.

Prolamines are deficient in lysine, and contain large amounts of proline; some prolamines are zein (from maize), gliadin (from wheat) and hordein (from barley).

(iv) *Glutelins.* These are insoluble in water or dilute salt solution, but are soluble in dilute acids and alkalis; they are coagulated by heat.

Some glutelins are glutenin (from wheat) and oryzenin (from rice).

(v) *Scleroproteins (albuminoids).* These are insoluble in water or salt solution, but are soluble in strong acids or alkalis.

Examples: keratin (from hair, hoof), fibroin (from silk); these are not attacked by enzymes.

Submembers of the scleroproteins are:

(*a*) *Collagens* (in skin, tendons and bones); these form gelatin (a water-soluble protein) when boiled with water. Collagens are attacked by pepsin or trypsin.

(*b*) *Elastins* (in tendons and arteries); these are not converted into gelatin, and are attacked slowly by trypsin.

(vi) *Basic proteins.* These are strongly basic, and fall into two groups.

(*a*) *Histones.* These are soluble in water or dilute acids, but are insoluble in dilute ammonia. They are not coagulated by heat, and contain large amounts of histidine and arginine. Histones are the proteins of the nucleic acids, hæmoglobin, etc.

that they are complex substances built up of amino-acids. It appears likely that all enzymes are proteins (see §12); many hormones are also proteins, *e.g.*, insulin.

Proteins may be coagulated, *i.e.*, precipitated irreversibly, by heat and by strong inorganic acids and bases, etc. When proteins are precipitated irreversibly, they are said to be *denatured*, but the chemical changes that occur in this process are still uncertain. Proteins may be precipitated by ethanol or concentrated solutions of ammonium sulphate or sodium chloride. In this case, the precipitation is reversible, *i.e.*, the precipitated proteins may be redissolved; thus they are not denatured by these reagents. Proteins are also precipitated by the salts of the heavy metals, *e.g.*, mercuric chloride, copper sulphate, etc., and they give many characteristic colour reactions with various reagents, *e.g.*,

(i) *Biuret reaction.* Addition of a very dilute solution of copper sulphate to an alkaline solution of a protein produces a red or violet colour.

(ii) *Millon's reaction.* When a solution of mercuric nitrate containing nitrous acid is added to a protein solution, a white precipitate is formed and slowly turns pink.

(iii) *Xanthoproteic reaction.* Proteins produce a yellow colour when treated with concentrated nitric acid.

Proteins are amphoteric, their behaviour as an anion or a cation depending on the *p*H of the solution. At some definite *p*H, characteristic for each protein, the solution contains equal amounts of anion and cation. In this condition the protein is said to be at its *isoelectric point*, and at this *p*H the protein has its least solubility, *i.e.*, it is most readily precipitated (*cf.* amino-acids, §4. C i). The osmotic pressure and viscosity of the protein solution are also a minimum at the isoelectric point. The amphoteric nature of proteins is due to the presence of a large number of free acidic and basic groups arising from the amino-acid units in the molecule. These groups can be titrated with alkali or acid, and by this means it has been possible to identify acidic and basic groups belonging to the various amino-acid units (see also §11).

The molecular weights of proteins have been determined by means of the ultracentrifuge, osmotic pressure measurements, X-ray diffraction, light scattering effects and by chemical analysis. Chemical methods are based on the estimation of a particular amino-acid, *e.g.*, casein contains cystine; hence the estimation of the percentage of this amino-acid and of sulphur will lead to the evaluation of the molecular weight of casein. The most reliable values of the molecular weights are those obtained by the ultracentrifuge method; the values recorded vary considerably for the

individual proteins, ranging from about 40,000 for egg albumin to about 5,000,000 for hæmocyanin.

§7. Classification of proteins. Several arbitrary classifications of the proteins are in use. One method is based mainly on physical properties, particularly solubility.

A. Simple proteins. These give only amino-acids on hydrolysis.

(i) *Albumins.* These are soluble in water (and in acids and alkalis), and are coagulated by heat. They are precipitated by saturating their solutions with ammonium sulphate.

Albumins are usually low or deficient in glycine; some albumins are serum albumin, egg albumin and lactalbumin.

(ii) *Globulins.* These are insoluble in water, but are soluble in dilute salt solution and in dilute solutions of strong inorganic acids and alkalis. They are precipitated by half saturating their solutions with ammonium sulphate, and they are coagulated by heat.

Globulins usually contain glycine; some typical globulins are serum globulin, tissue globulin and vegetable globulin.

(iii) *Prolamines.* These are insoluble in water or salt solution, but are soluble in dilute acids and alkalis, and in 70–90 per cent. ethanol.

Prolamines are deficient in lysine, and contain large amounts of proline; some prolamines are zein (from maize), gliadin (from wheat) and hordein (from barley).

(iv) *Glutelins.* These are insoluble in water or dilute salt solution, but are soluble in dilute acids and alkalis; they are coagulated by heat.

Some glutelins are glutenin (from wheat) and oryzenin (from rice).

(v) *Scleroproteins* (*albuminoids*). These are insoluble in water or salt solution, but are soluble in strong acids or alkalis.

Examples: keratin (from hair, hoof), fibroin (from silk); these are not attacked by enzymes.

Submembers of the scleroproteins are:

(*a*) *Collagens* (in skin, tendons and bones); these form gelatin (a water-soluble protein) when boiled with water. Collagens are attacked by pepsin or trypsin.

(*b*) *Elastins* (in tendons and arteries); these are not converted into gelatin, and are attacked slowly by trypsin.

(vi) *Basic proteins.* These are strongly basic, and fall into two groups.

(*a*) *Histones.* These are soluble in water or dilute acids, but are insoluble in dilute ammonia. They are not coagulated by heat, and contain large amounts of histidine and arginine. Histones are the proteins of the nucleic acids, hæmoglobin, etc.

(b) *Protamines.* These are more basic than the histones and have a simpler structure. They are soluble in water, dilute acids and dilute ammonia; they are not coagulated by heat, and are precipitated from solution by ethanol. They contain large amounts of arginine, and occur in various nucleic acids.

B. Conjugated proteins are proteins which contain a non-protein group (*i.e.*, a compound not containing amino-acid residues) attached to the protein part. The non-protein group is known as the *prosthetic group*, and it may be separated from the protein part by careful hydrolysis.

(i) *Nucleoproteins.* The prosthetic group is a nucleic acid.

(ii) *Chromoproteins.* These are characterised by the presence of a metal, *e.g.*, iron, magnesium, copper, manganese, cobalt, etc. Chromoproteins may also contain a coloured prosthetic group. Examples: chlorophyll and hæmoglobin.

(iii) *Glycoproteins.* In these the prosthetic group contains a carbohydrate or a derivative of the carbohydrates.

(iv) *Phosphoproteins.* These are conjugated proteins in which the prosthetic group contains phosphoric acid in some form other than in the nucleic acids or in the lipoproteins.

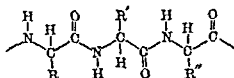
(v) *Lipoproteins.* In these the prosthetic group is lecithin, kephalin, etc.

(vi) *Metalloproteins.* These are heavy metal-protein complexes; all the heavy metals can form complex ions with proteins, *e.g.*, calcium caseinate occurs in blood.

C. Derived proteins are degradation products obtained by the action of acids, alkalis or enzymes on proteins.

Protein ⟶ *Denatured proteins*; insoluble proteins formed by the action of heat, etc., on proteins.
↓
Primary proteoses (metaproteins); insoluble in water or dilute salt solution, but are soluble in acids or alkalis. They are precipitated by half-saturation *with ammonium sulphate.*
↓
Secondary proteoses; soluble in water, not coagulated by heat, and are precipitated by saturatino with ammonium sulphate.
↓
Peptones
↓
Polypeptides
↓
Simple peptides
(*Peptones*, *Polypeptides*, *Simple peptides*: These are soluble in water, not coagulated by heat, and are not precipitated by saturation with ammonium sulphate.)
↓
Amino-acids

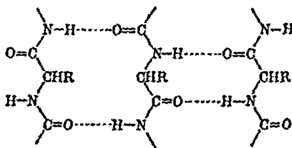
§8. Structure of the proteins. Proteins are hydrolysed by acids, alkalis or by suitable enzymes to a mixture of amino-acids. About twenty-five acids have been definitely isolated; all or only some of these acids may be present in a given protein, and their proportions vary from protein to protein. Fischer (1902) and Hofmeister (1902) suggested that amino-acids in proteins are joined in a *linear* fashion by *peptide* linkages, *i.e.*, by the —CO·NH— group, the carboxyl group of one amino-acid molecule forming an amide by combination with the amino-group of the next amino-acid molecule, etc. When a relatively small number of amino-acids are linked together (as amides), the resulting molecule is called a *peptide*. When a relatively large number of amino-acid residues are present in the molecule, then that compound is called a *polypeptide*. Proteins are far more complex than the polypeptides. Thus, on this basis, a protein molecule may be represented as a *linear polymer* of amino-acid molecules.



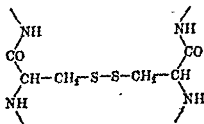
The examination of the infra-red absorption spectra of various synthetic polypeptides has shown the presence of the peptide (*i.e.*, amide) link, and that these links are at positions expected for them. Furthermore, it has been shown that proteins of the keratin type have bands characteristic of the peptide link (Darmon *et al.*, 1947).

Since some amino-acids contain two amino or two carboxyl groups, it is therefore possible to have *free* amino and carboxyl groups at various positions along the chain, *i.e.*, the group R may contain a free amino or carboxyl group. Since the hydrolysis of certain proteins leads to the formation of ammonia, it has been concluded that in addition to free amino and carboxyl groups, *there are also some carbonamide groups, —$CONH_2$. X-ray analysis* has confirmed the existence of these polypeptide chains. Furthermore, these chains are arranged in a three-dimensional lattice, the chains being held together, to a large extent, by hydrogen bonds. Chibnall (1942) has suggested that protein molecules may be composed of several *separate* long parallel chains. In this case the chains are held together by *covalent bonds*; it is possible for the carboxyl group on one chain to form links with an amino-group on another chain (R may contain either of these groups).

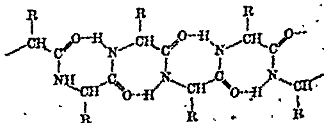
On the other hand, when the protein contains cystine, there is



a possibility that the chains are *cross-linked via* sulphur (*cf.* vulcanisation of rubber, §33a. VIII).



Proteins have been found to be of two types, *fibrous* and *globular*. In fibrous proteins the polypeptide chains are extended. In some cases, however, the chains are apparently " coiled ", and these may be extended by the application of a force. The nature of the coiled structure is uncertain, but two configurations have been proposed which agree reasonably well with information obtained from infrared spectra, X-ray data, bond lengths and bond energies. According to Ambrose *et al.* (1949), the polypeptide chain is folded into a series of seve ... the chain being stabilised by hy ... fibre, a number of these folded ... e).



On the other hand, Pauling *et al.* (1951) have proposed a coiled chain in the form of a helix containing either 3·7 or 5·1 acid residues.

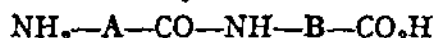
The folded (coiled) form of a fibrous protein is known as the *α-form*, and the extended as the *β-form*. Elliott *et al.* (1951, 1953)

have observed that the frequency of the CO stretching mode in synthetic polypeptides and natural proteins depends on the configuration of the polypeptide chain. Thus this offers a means of distinguishing between the α- and β-forms.

The globular (corpuscular) proteins are more compact than the fibrous proteins, but their shape is not spherical; *e.g.*, X-ray studies have shown that hæmoglobin has a cylindrical shape. The arrangement of the polypeptide chains in these globular proteins is still uncertain. An interesting point in connection with globular proteins is that infra-red methods may be used to detect the presence of carboxylate groups in them at the isoelectric point (Ehrlich *et al.*, 1954).

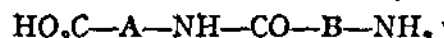
Of the two types of proteins, it is only the globular which have been obtained crystalline; the fibrous proteins lack the characteristics necessary for crystallisation. It appears that all protein crystals grown from solution contain solvent, the removal of which causes the protein to become less crystalline. The solvent has been shown to be interstitial and not "solvent of crystallisation".

One other point about the nature of these polypeptide chains will now be mentioned briefly. Let us consider a dipeptide composed of two *different* amino-acids, A and B. These may be combined in two different ways:



$$NH_2{-}A{-}CO{-}NH{-}B{-}CO_2H$$

and

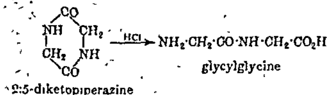


$$HO_2C{-}A{-}NH{-}CO{-}B{-}NH_2$$

[illegible] be combined in six different ways. [illegible] ds, there will be *n*! different com- [illegible] he naturally occurring amino-acids (excluding glycine) been all of the L-series, the total number of possible combinations would have been very much larger still. It is therefore of great interest to ascertain the "order" in which amino-acids are combined in proteins. Some progress has been made in this direction (see §11).

§9. Synthesis of polypeptides. Various methods have been introduced, *e.g.*,

(i) The partial hydrolysis of a diketopiperazine with hydrochloric acid gives a dipeptide (Fischer, 1901), *e.g.*,



Glycylglycine was the first peptide to be synthesised. The method is very limited in application, since only dipeptides may be prepared. If a "mixed" diketopiperazine is used, then hydrolysis can proceed in two different ways; the nature of the product depends on the hydrolysing agent used.

(ii) The methyl esters of di- and tri-peptides tend to eliminate methanol to form a higher peptide (Fischer, 1901).





$$2NH_2{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2CH_3 \longrightarrow NH_2{\cdot}CH_2{\cdot}CO{\cdot}(NH{\cdot}CH_2{\cdot}CO)_2{\cdot}NH{\cdot}CH_2{\cdot}CO_2CH_3$$

By means of this reaction, Frankel *et al.* (1942) prepared polypeptides containing up to 110 glycyl units.

(iii) When two different amino-acids are joined to form a dipeptide, two possibilities occur (*cf.* §8); thus, if glycine and alanine are linked together, the two possibilities are:



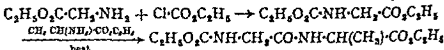
$$NH_2{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH(CH_3){\cdot}CO_2H$$

and



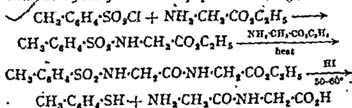
$$NH_2{\cdot}CH(CH_3){\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2H$$

In order to condense the two amino-acids in a known manner, Fischer (1901, 1903) "blocked" the amino-group of one molecule by first reacting that compound with ethyl chloroformate; thus:



$$C_2H_5O_2C{\cdot}CH_2{\cdot}NH_2 + Cl{\cdot}CO_2C_2H_5 \longrightarrow C_2H_5O_2C{\cdot}NH{\cdot}CH_2{\cdot}CO_2C_2H_5 \xrightarrow[\text{heat}]{CH_3{\cdot}CH(NH_2){\cdot}CO_2C_2H_5} C_2H_5O_2C{\cdot}NH{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH(CH_3){\cdot}CO_2C_2H_5$$

Hence by starting with glycylglycine, the diglycylalanine derivative may be prepared. The difficulty with this method, however, is that it is not possible to remove the *N*-carbethoxyl group by hydrolysis without also hydrolysing the peptide link. This difficulty was overcome by Fischer (1915) by using *p*-toluenesulphonyl chloride as the "blocking agent" instead of ethyl chloroformate; the former group can be removed (as the thiophenol) by warming with hydriodic acid, without hydrolysis of the peptide link, *e.g.*,

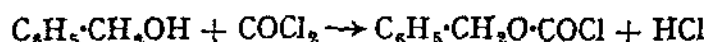


$$CH_3{\cdot}C_6H_4{\cdot}SO_2Cl + NH_2{\cdot}CH_2{\cdot}CO_2C_2H_5 \longrightarrow CH_3{\cdot}C_6H_4{\cdot}SO_2{\cdot}NH{\cdot}CH_2{\cdot}CO_2C_2H_5 \xrightarrow[\text{heat}]{NH_2{\cdot}CH_2{\cdot}CO_2C_2H_5}$$

$$CH_3{\cdot}C_6H_4{\cdot}SO_2{\cdot}NH{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2C_2H_5 \xrightarrow[50\text{-}60^\circ]{HI} CH_3{\cdot}C_6H_4{\cdot}SH + NH_2{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2H$$

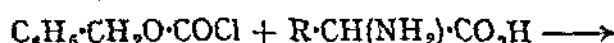
Bergmann (1932) found that benzyl chloroformate is a very good blocking agent, and its application is much wider than that of *p*-toluenesulphonyl chloride. Benzyl chloroformate is readily prepared by the action of carbonyl chloride on benzyl alcohol in toluene

solution (benzyl chloroformate is also known as carbobenzyloxy or benzyloxycarbonyl chloride):



$$C_6H_5{\cdot}CH_2OH + COCl_2 \rightarrow C_6H_5{\cdot}CH_2O{\cdot}COCl + HCl$$

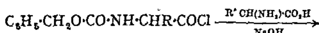
The procedure is then as follows:



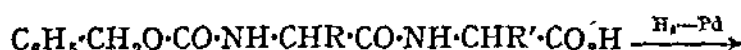
$$C_6H_5{\cdot}CH_2O{\cdot}COCl + R{\cdot}CH(NH_2){\cdot}CO_2H \longrightarrow$$



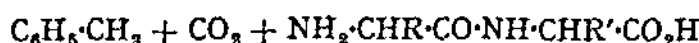
$$C_6H_5{\cdot}CH_2O{\cdot}CO{\cdot}NH{\cdot}CHR{\cdot}CO_2H \xrightarrow{PCl_5}$$



$$C_6H_5{\cdot}CH_2O{\cdot}CO{\cdot}NH{\cdot}CHR{\cdot}COCl \xrightarrow[NaOH]{R'CH(NH_2){\cdot}CO_2H}$$



$$C_6H_5{\cdot}CH_2O{\cdot}CO{\cdot}NH{\cdot}CHR{\cdot}CO{\cdot}NH{\cdot}CHR'{\cdot}CO_2H \xrightarrow{H_2-Pd}$$



$$C_6H_5{\cdot}CH_3 + CO_2 + NH_2{\cdot}CHR{\cdot}CO{\cdot}NH{\cdot}CHR'{\cdot}CO_2H$$

If the amino-acid contains sulphur, then catalytic reduction cannot be used, since the sulphur poisons the catalyst; the removal of the blocking group, however, may be successfully accomplished by means of phosphonium iodide or sodium in liquid ammonia.

A more recent and convenient method of removing the benzyloxycarbonyl group is to treat the derivative with hydrogen bromide in acetic acid or nitromethane (Ben-Ishai *et al.*, 1952; Anderson *et al.*, 1952):



$$C_6H_5{\cdot}CH_2O{\cdot}CO{\cdot}NH{\cdot}CHR{\cdot}CO{\cdot}NH{\cdot}CHR'{\cdot}CO_2H \xrightarrow{HBr}$$

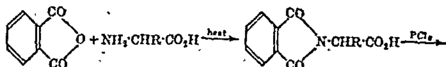


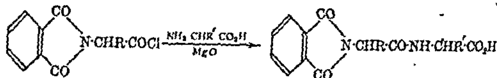
$$C_6H_5{\cdot}CH_2Br + CO_2 + \overset{-}{Br}\overset{+}{N}H_3{\cdot}CHR{\cdot}CO{\cdot}NH{\cdot}CHR'{\cdot}CO_2H$$

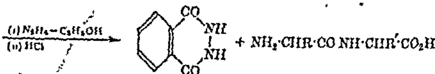
Weisblat *et al.* (1953) have also shown that the *p*-toluenesulphonyl group can be removed by means of hydrogen bromide in acetic acid containing phenol.

Stevens *et al.* (1950) have used allyl chloroformate instead of benzyl chloroformate, and then removed the carboallyloxy-group by means of sodium in liquid ammonia.

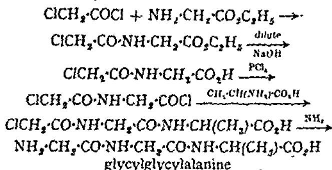
Sheehan *et al.* (1949) have used the following method for blocking the amino-group of one amino-acid residue [*cf.* Gabriel's phthalimide synthesis, §2 (*ib*)].







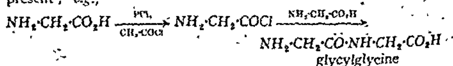
(iv) Polypeptides may be synthesised by combining an α-halogenoacid chloride with an amino-acid ester and then proceeding as follows (Fischer, 1903).



$$ClCH_2{\cdot}COCl + NH_2{\cdot}CH_2{\cdot}CO_2C_2H_5 \longrightarrow$$
$$ClCH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2C_2H_5 \xrightarrow[NaOH]{dilute}$$
$$ClCH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2H \xrightarrow{PCl_5}$$
$$ClCH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}COCl \xrightarrow{CH_3{\cdot}CH(NH_2){\cdot}CO_2H}$$
$$ClCH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH(CH_3){\cdot}CO_2H \xrightarrow{NH_3}$$
$$NH_2{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH(CH_3){\cdot}CO_2H$$

glycylglycylalanine

(v) A variation of the previous method is to convert an amino-acid into its corresponding acid chloride by means of phosphorus pentachloride in acetyl chloride, and then to treat the acid chloride with another molecule of an amino-acid (Fischer, 1907). In the formation of the acid chloride, hydrogen chloride is also produced, and this combines with the amino-group to form the group $\bar{C}l\{H_3\overset{+}{N}{\cdot}CHR{\cdot}$, which is *not* acetylated by the acetyl chloride present; *e.g.*,



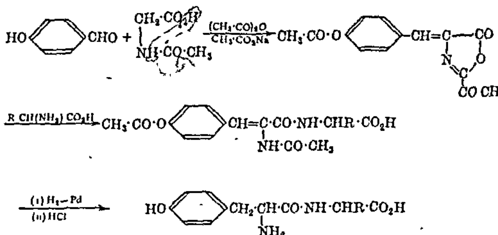
$$NH_2{\cdot}CH_2{\cdot}CO_2H \xrightarrow[CH_3{\cdot}COCl]{PCl_5} NH_2{\cdot}CH_2{\cdot}COCl \xrightarrow{NH_2{\cdot}CH_2{\cdot}CO_2H} NH_2{\cdot}CH_2{\cdot}CO{\cdot}NH{\cdot}CH_2{\cdot}CO_2H$$

glycylglycine

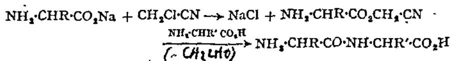
By this means Fischer (1907) succeeded in synthesising an octadecapeptide (of molecular weight 1213), and Abderhalden (1916) synthesised a nonadecapeptide (of molecular weight 1326).

(vi) The above methods involving the intermediate formation of an acid chloride cannot be applied to hydroxyamino- and di-amino-acids, since these acids react with phosphorus pentachloride in a complicated fashion and do not give the desired halogen compounds.

In such cases Bergmann (1926) successfully applied the azlactone synthesis, *e.g.*,



(vii) A very recent method of building up peptides is that of Schwyzer *et al.* (1955), this method involves the use of chloroacetonitrile as follows:

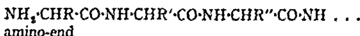


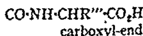
$$NH_2\cdot CHR\cdot CO_2Na + CH_2Cl\cdot CN \longrightarrow NaCl + NH_2\cdot CHR\cdot CO_2CH_2\cdot CN \xrightarrow{NH_2\cdot CHR'\cdot CO_2H} NH_2\cdot CHR\cdot CO\cdot NH\cdot CHR'\cdot CO_2H$$

As we have seen (§4), all the amino-acids except glycine contain at least one asymmetric carbon atom. Furthermore, the α-acylamino-acids are readily racemised, and hence a very important point about the syntheses described above is that racemisation will occur during the syntheses. The actual extent of racemisation depends on the nature of the acyl group and the type of condensation used. According to Boissonas *et al.* (1955), the benzyloxycarbonyl group gives very resistant derivatives (to racemisation) and is therefore the best one to use.

§10. Properties of the polypeptides. The polypeptides are solids which usually decompose when heated to 200–300°. They are soluble in water, but are insoluble in ethanol, and have a bitter taste similar to that of the proteins. They are hydrolysed by acids, alkalis and enzymes, and they very closely resemble the polypeptides actually obtained by the partial hydrolysis of proteins. Polypeptides (synthetic) also give the biuret test. Many peptides have been found as the products of metabolism of micro-organisms.

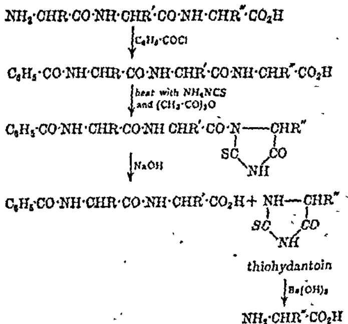
§11. Degradation of the polypeptides. It has already been pointed out that a necessary requirement for the elucidation of the structure of proteins is a knowledge of the "order" of the amino-acid residues in the molecule (§8). Chemical methods have been introduced whereby the *terminal* amino-acid residue of a polypeptide may be removed in a stepwise fashion. Consideration of the following structure of a polypeptide shows that the two ends of the molecule are not alike; the end on the left-hand side is known as the "amino-end", and that on the right-hand side as the "carboxyl-end".





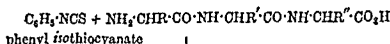
Methods have been introduced for degrading either the carboxyl-end or the amino-end of the polypeptide chain, *e.g.*,

Carboxyl-end degradation. The following method is due to Schlack and Kumpf (1926).

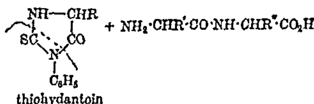


Thus the terminal amino-acid can be identified, and the process can now be repeated on the degraded peptide.

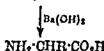
Amino-end degradation. The following method is due to Edman (1950).







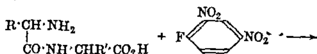
thiohydantoin

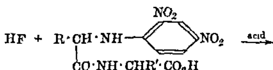


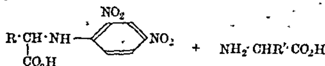
Thus the terminal amino-acid can be identified, and the process can now be repeated on the degraded peptide.

More recently, Asai *et al.* (1955) have investigated the infra-red spectra of polypeptides and have shown that certain bands depend largely on the *sequence* of the amino-acids in the chain. These authors have concluded that the crystalline part of silk fibroin contains glycine and alanine residues arranged alternately.

Another interesting point about the structure of polypeptides is the nature of the amino-acids in the chain which have free amino groups (*cf.* §6). Sanger (1945) has developed the "DNP" method for solving this problem. He showed that 1-fluoro-2·4-dinitrobenzene reacts readily only with amino groups and forms derivatives which are stable to acids, *e.g.*,







Thus, when a peptide is first treated with the reagent and then the product hydrolysed with acid, a number of amino-acids will be obtained as their dinitrophenyl derivatives (which can be separated by chromatography; *cf.* §3).

ENZYMES

§12. **General nature of enzymes.** Enzymes are biological catalysts which bring about chemical reactions in living cells. They are produced by the living organism, and are usually present in only very small amounts in the various cells (about 0·01 per cent.). They can also exhibit their activity even when they have been extracted from their source. The enzymes are all organic compounds, and a number of them have been obtained in a crystalline form. Those so far [illegible] high molecular weig[illegible] some are yellow, blue, green or greenish-brown; most are soluble in water or dilute salt solution. Some enzymes are purely protein in nature, but many contain a prosthetic group (see §7. B) which has a relatively low molecular weight. The prosthetic group of some enzymes is readily separated (*e.g.*, by dialysis) from the protein part and the latter, in this condition, is known as an *apoenzyme*, *e.g.*, peroxidase is composed of hæmatin (prosthetic group; see §2. XIX) linked with the protein (the apoenzyme). The prosthetic group is often referred to as the co-enzyme (when dealing with enzymes); both parts must be present for the "enzyme" to act.

§13. **Nomenclature.** The systematic method of naming enzymes is to add the suffix *ase* to the name of the *substrate*, *i.e.*, the substance being acted upon, *e.g.*, esterase acts on esters, amylase on starch (amylum), protease on proteins, urease on urea, etc. Some enzymes, however, have retained their trivial names, *e.g.*, emulsin, pepsin, trypsin, etc. Names are also used for *particular* enzymes, *e.g.*, urease, amylase, or as *general* names for *groups* of enzymes, *e.g.*, esterases, proteases, etc. Enzymes of various species are quite often similar, and the reactions catalysed by them are identical. Even so, it does not necessarily follow that these enzymes are identical chemically, *e.g.*, amylases from different sources have different *p*H optima (see below).

§14. **Classification of enzymes.** Enzymes are usually classified on the type of reaction which they catalyse. There are two main groups:

(i) *Hydrolytic enzymes.* These bring about hydrolysis, *e.g.*, proteases (proteins), lipases (esters), carbohydrases (carbohydrates), etc.

(ii) *Oxidative enzymes.* Most oxidative enzymes function by transferring hydrogen from the substrate (or a modified form, *e.g.*, a hydrated form) to themselves, *i e*, they behave as hydrogen acceptors. These enzymes are known as *dehydrogenases.* There are also a few enzymes which oxidise the substrate directly with molecular oxygen ; these are known as *oxidases, e g., ascorbic acid oxidase* catalyses the oxidation of ascorbic acid to dehydroascorbic acid by molecular oxygen (*cf* §11. VII).

Some other types of enzymes are isomerising enzymes, transferring enzymes (*e g.*, transaminases catalyse the transfer of an amino group of an amino-acid to a keto group of a keto acid), and "splitting enzymes" (*e g.*, decarboxylases catalyse decarboxylation).

§15. Conditions for enzyme action. A number of factors influence enzyme activity: the concentration of the enzyme, the concentration of the substrate, the *p*H of the solution, and the temperature. The optimum conditions for a particular enzyme must be found experimentally. The optimum *p*H varies considerably for individual enzymes, and for a given enzyme, with the nature of the substrate. The optimum temperature for animal enzymes is usually between 40° and 50°, and that for plant enzymes 50° and 60°. Most enzymes are irreversibly destroyed when heated above 70–80°.

Many enzymes have been shown to be reversible in their action, *i.e.*, they can both degrade and synthesise. The optimum conditions, however, for degradation are very often totally different from those for synthesis. Furthermore, it does not follow that synthesis in the organism is effected by the same enzyme which produces degradation, *e.g.*, urea is hydrolysed by urease in plants, but is formed in animals by the action of arginase on the amino-acid arginine.

§16. Specificity of enzyme action. One of the most characteristic properties of enzymes is their specificity of action. This specificity may be manifested in one of three ways:

(i) Specificity for a particular reaction or a particular type of reaction, *e.g.*, urease will hydrolyse only urea ; esterases hydrolyse only esters. Enzymes may also be specific within a group, *e.g.*, phosphatases (a group of esterases) only hydrolyse esters in which the acid component is phosphoric acid.

(ii) Many enzymes exhibit a *relative specificity*, *e.g.*, esterase

although hydrolysing all esters, hydrolyse the various esters at *different* speeds; pepsin hydrolyses the peptide link, but is most active for those links in which, among other things, the amino group belongs to an aromatic amino-acid and the carboxyl group is one of a dicarboxylic amino-acid.

(iii) Many enzymes are *stereospecific*, *e.g.*, maltase hydrolyses α-glycosides but not β-glycosides, whereas emulsin hydrolyses the latter but not the former (*cf.* §3. VII).

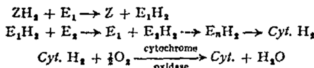
It should be noted, however, that a given enzyme can exhibit more than one of these specificities, *e.g.*, esterases, while hydrolysing only esters, may also hydrolyse one enantiomorph (of an optically active ester) more rapidly than the other.

Another point of interest here is that the general type of reaction catalysed by an enzyme depends on the nature of the prosthetic group, and the specificity of the enzyme depends on the nature of the apoenzyme (protein).

§17. Mechanism of enzyme action. According to one view, enzymes initiate the reaction, but according to another view, the reaction catalysed by an enzyme is capable of proceeding at a very slow rate in the absence of the enzyme (*cf.* the theories of catalysis).

The details of the mechanism of the catalysis effected by enzymes are still not certain. A highly favoured theory is that the enzyme passes through a transition state by combination with its substrate, and then the enzyme is regenerated with the simultaneous formation of the products (*cf.* the transition state, §3b. III). A number of these transition states have been shown to exist from, *e.g.*, spectroscopic evidence; during the reaction the absorption spectrum of the enzyme is altered. It is also believed that the protein part of the enzyme has an "active centre", and it is this which combines with the substrate. Assuming this be the case, it is now necessary to explain why neither the protein part of the enzyme nor the prosthetic group can act separately, but both must be present (apparently in combination). The answer to this question has been given in certain cases, *e.g.*, with dehydrogenases it has been suggested that the function of the prosthetic group is to act as a hydrogen acceptor, and that the function of the protein part of the enzyme is to "facilitate" the transfer of the hydrogen from the activated complex. It appears that the usual dehydrogenase action occurs in a number of steps involving different enzymes acting as hydrogen acceptors. Thus each enzyme undergoes reversible reduction and oxidation, and finally the last step is catalysed by cytochrome which is reduced, and this is reoxidised to cytochrome (and water) by

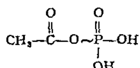
molecular oxygen by means of *cytochrome oxidase*. The sequence may therefore be represented:



$$ZH_2 + E_1 \longrightarrow Z + E_1H_2$$
$$E_1H_2 + E_2 \longrightarrow E_1 + E_2H_2 \dashrightarrow E_nH_2 \longrightarrow \textit{Cyt.}\ H_2$$
$$\textit{Cyt.}\ H_2 + \tfrac{1}{2}O_2 \xrightarrow[\text{oxidase}]{\text{cytochrome}} \textit{Cyt.} + H_2O$$

Most of the dehydrogenases contain a prosthetic group (which is the hydrogen acceptor). Thus a number of vitamins (§1. XVII) function as part of prosthetic groups, *e.g.*, pyridino-enzymes (pyridine nucleus), flavo-enzymes (riboflavin), etc. On the other hand, cytochrome and catalase are hæm-containing enzymes, and ascorbic acid oxidase and phenolase are copper protein enzymes.

When small molecules are converted into large molecules containing more energy than the units from which they were built, then energy must be supplied to bring about these syntheses. Enzymes are involved in these syntheses, and it is believed that certain organic compounds contain *energy-rich phosphate bonds*, and when dephosphorylation occurs energy is liberated, *e.g.*, acetyl phosphate contains such a bond (the symbol ~ is used to represent an energy-rich bond):



Many enzymes are inactive unless an *activator* is present. The inactive enzyme is known as a *zymogen*, and the activator as a *kinase* (if this is inorganic), *e.g.*, trypsinogen (the zymogen) together with enterokinase (the kinase) forms the enzyme trypsin. Some activators may be metallic or non-metallic, *e.g.*, salivary amylase requires chloride ions for activity. There appears to be no sharp distinction between an activator and a co-enzyme. Some authors describe activators as those compounds which are *non-specific*, *e.g.*, chloride ions, and co-enzymes as compounds which are *specific*.

Many substances may behave as *inhibitors*, *i.e.*, in their presence the enzyme fails to act; *e.g.*, saccharase is inactivated by copper ions (*cf.* "poisons" in catalysis). Sometimes purely physical means may inactivate an enzyme, *e.g.*, crystalline pepsin is inactivated by sound waves with a frequency of 9 kilocycles per second.

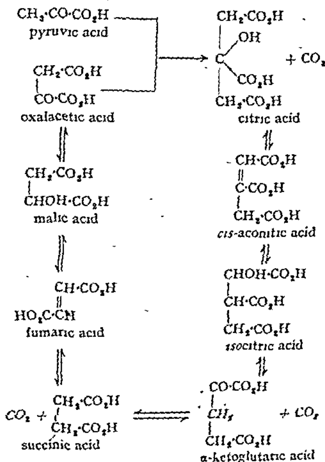
§18. Biosynthesis of amino-acids and proteins. First let us consider the *Krebs cycle* (1937). This is also known as the *citric acid cycle* and is the scheme proposed for the biological oxidation

of hexoses to carbon dioxide and water. The first step is the conversion of a hexose molecule into two molecules of pyruvic acid; this occurs *via* the formation of phosphoglyceraldehyde (*cf.* §23a. VII). The pyruvic acid combines with carbon dioxide to form oxalacetic acid:



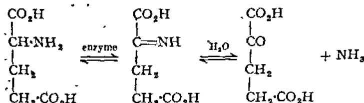
$$CH_3{\cdot}CO{\cdot}CO_2H + CO_2 \rightleftharpoons CO_2H{\cdot}CH_2{\cdot}CO{\cdot}CO_2H$$

The Krebs cycle may then be written as follows (the various enzymes involved and mechanisms are not shown):

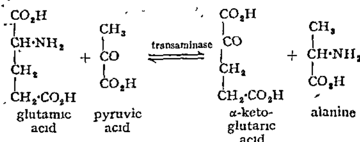


Amino-acids can be deaminated to keto-acids, and in addition to the general (amino-acid) dehydrogenases, there is a specific *glycine dehydrogenase* and a specific *glutamic dehydrogenase*. The case of glutamic acid is extremely important, since there is much evidence to show that this acid plays a vital part in the metabolism of amino-acids. Furthermore, it appears that the conversion of

glutamic acid into α-ketoglutaric acid is the only *reversible* reaction in the oxidative deamination of amino-acids.



Keto-acids produced by deamination of amino-acids may undergo further transformations, one being their conversion into amino-acids. This, however, occurs by the process of transamination under the influence of *transaminases*, *e g*,



We have already seen (§32a. VIII) how various keto-acids could be synthesised in the organism. Thus, with the formation of α-ketoglutaric acid from the break-down of carbohydrates, its direct amination to glutamic acid, and the latter now capable of aminating other keto-acids by transamination, the cycle of events is set up for the biosynthesis of amino-acids in general. A point to be noted in this connection is that some amino-acids are essential (§1), *e.g.*, man cannot synthesise the benzene ring.

A very interesting problem related to the biosynthesis of amino-acids is the work of Miller (1953, 1955). This author subjected a mixture of methane, ammonia, hydrogen and water vapour (which possibly made up the atmosphere of the Earth in its early stages) to spark and silent discharges. Analysis of the gases showed that the initial gases were present and, in addition, carbon monoxide, carbon dioxide, and nitrogen. The solid product was analysed by means of paper chromatography, and the following amino-acids were identified: glycine, sarcosine (N-methylglycine), [illegible] alanine, β-alanine, D- and [illegible]-aminobutyric acid, and [illegible] isobutyric acid. Many other amino-acids (unidentified) were formed, as well as formic, acetic, propionic, glycollic [illegible] acids.

Bahadur (1954), on the other hand, has synthesised amino-acids by exposing a solution of paraformaldehyde and potassium nitrate to bright sunlight

Finally, let us consider the biosynthesis of the proteins from amino-acids. Many workers have concluded that there are no intermediates, *i.e.*, protein synthesis is an "all-at-once" assembly of amino-acids. On the other hand, other workers have concluded that intermediates are formed, but these are so poorly defined or are so transient that they cannot be characterised. Steinberg *et al* (1951–), using amino-acids labelled with ^{14}C, have shown that their results are compatible with the step-wise mechanism through intermediates.

Reading References

Schmidt, *The Chemistry of the Amino-Acids and Proteins*, Thomas (1943, 2nd ed.)

Sahyum (Ed.), *Outline of the Amino-Acids and Proteins*, Reinhold (1948, 2nd ed.).

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol. II (1943, 2nd ed). Ch. 14. Natural Amino-Acids

Rodd (Ed), *Chemistry of Carbon Compounds*, Elsevier. Vol. 1B (1952). Ch. 22. Proteins.

Greenberg (Ed), *Amino-Acids and Proteins*, Thomas (1951).

Sidgwick, *The Organic Chemistry of Nitrogen*, Oxford Press (New Edition by Taylor and Baker, 1942). Ch. IV. Amino-Acids.

Springall, *The Structural Chemistry of Proteins*, Academic Press (1950).

Advances in Protein Chemistry, Academic Press (1944–).

Editorial Report on Nomenclature, *J.C.S.*, **1951**, 3522. Rules for the Nomenclature of Natural Amino-Acids and Related Substances.

Organic Reactions, Wiley. Vol III (1946). Ch. 5. Azlactones.

Baltazzi, The Chemistry of 5-Oxazolones, *Quart. Reviews (Chem. Soc.)*, **1955, 9,** 150.

Synge, Naturally Occurring Peptides, *Quart. Reviews (Chem. Soc.)*, **1949, 3,** 245.

Khorana, Structural Investigation of Peptides and Proteins, *Quart. Reviews (Chem. Soc.)*, **1952, 6,** 340.

Heilbron, Concerning Amino-Acids, Peptides, and Purines, *J.C.S.*, **1949,** 2099.

Dodds, Symposium on Amino-Acids and Protein Hydrolysates: Introductory Address, *Chem. and Ind.*, **1950,** 135.

Atkinson *et al.*, Recent Developments in the Synthesis of the Amino-Acids, *Chem. and Ind.*, **1951,** 118.

Shapiro, The Synthesis of Peptides, *Chem. and Ind.*, **1952,** 1119.

Tristram, Amino-Acid Analysis, *Chem. and Ind.*, **1950,** S868.

Asimov, Potentialities of Protein Isomerism, *J. Chem. Educ.*, **1954, 31,** 125.

Springall and Law, Peptides: Methods of Synthesis and Terminal-Residue Studies, *Quart. Reviews (Chem. Soc.)*, **1956, 10,** 230.

Kenner, Recent Progress in the Chemistry of Peptides, *J C S.*, 1956, 3689.
Steinberg *et al.*, Kinetic Aspects of Assembly and Degradation of Proteins, *Science*, 1956, 124, 389
Sumner and Somers, *Chemistry and Methods of Enzymes*, Academic Press (1947, 2nd ed).
Sumner, Enzymes, The Basis of Life, *J. Chem. Educ*, 1952, 29, 114.
Avison and Hawkins, The Role of Phosphoric Esters in Biological Reactions, *Quart. Reviews (Chem. Soc)*, 1951, 5, 171
Hoffmann-Ostenof, Suggestions for a More Rational Classification and Nomenclature of Enzymes, *Advances in Enzymology*, Interscience Publishers, 1953, 14, 219.
Klyne (Ed.), *Progress in Stereochemistry*, Butterworth (1954).
(i) Ch. 7. The Stereo
Weight. (ii) Ch. 8.
Newer Methods of Prepar
lishers (1948) The U
tions for Preparative Purposes (pp. 159–196).
Challenger, Biological Methylation, *Quart. Reviews (Chem. Soc.)*, 1955, 9, 255.
Miller, Production of Some Organic Compounds under Possible Primitive Earth Conditions, *J. Amer. Chem. Soc*, 1955, 77, 2351.
Downes, *The Chemistry of Living Cells*, Longmans, Green (1955).
Dixon and Webb, *Enzymes*, Longmans, Green (1958).

CHAPTER XIV

ALKALOIDS

§1. Definition of an alkaloid. *Originally the name* alkaloi (which means alkali-like) was given to *all organic bases isolate from plants.* This definition covers an extraordinary wide variet of compounds, and as the study of "alkaloids" progressed, so th definition changed. Königs (1880) suggested that alkaloids shoul be defined as naturally occurring organic bases which contain a pyridine ring. This definition, however, embraces only a limited number of compounds, and so the definition was again modified a little later by Ladenburg, who proposed to define alkaloids as natural plant compounds having a basic character and containing at least one nitrogen atom in a heterocyclic ring. Ladenburg's definition excludes any synthetic compounds and any compounds obtained from animal sources. One must admit that even to-day it is still difficult to define an alkaloid. The term is generally limited to organic bases formed in plants. Not all authors do this, and so they specify those alkaloids obtained from plants as *plant alkaloids (or vegetable alkaloids)*. *On the whole, alkaloids are very poisonous,* but are used medicinally in very small quantities. Thus we find that the basic properties, physiological action, and plant origin [illegible] which define plant alkaloids. *Even so, the* [illegible] which possess [illegible] the heading of alkaloids (some purines are also obtained from animal sources).

It is interesting to note in this connection that Sertürner (1806) isolated a basic compound from opium. Up to that time it was believed that plants produced only acids or neutral compounds.

§2. Extraction of alkaloids. In general, the plant is finely powdered and extracted with ethanol. The solvent is then distilled off, and the residue treated with dilute inorganic acids, whereupon the bases are extracted as their soluble salts. The free bases are liberated by the addition of sodium carbonate and extracted with various solvents, *e.g.*, ether, chloroform, etc. The mixtures of bases thus obtained are then separated by various methods into the individual compounds.

§3. General properties. The alkaloids are usually colourless, crystalline, non-volatile solids which are insoluble in water, but are

soluble in ethanol, ether, chloroform, etc. Some alkaloids are liquids which are soluble in water, *e.g.*, coniine and nicotine, and a few are coloured, *e.g.*, berberine is yellow. Most alkaloids have a bitter taste and are *optically active. They are generally tertiary nitrogen* compounds and contain one or two nitrogen atoms usually in the tertiary state in a ring system; most of the alkaloids also contain oxygen. The optically active alkaloids are very useful for resolving racemic acids. The alkaloids form insoluble precipitates with solutions of phosphotungstic acid, phosphomolybdic acid, picric acid, potassium mercuri-iodide, etc. Many of these precipitates have definite crystalline shapes and so may be used to help in the identification of an alkaloid.

§4. General methods for determining structure.

(i) After a pure specimen has been obtained it is subjected to *qualitative analysis (invariably the alkaloid contains (carbon),* hydrogen and nitrogen; most alkaloids also contain oxygen). This is then followed by quantitative analysis and thus the empirical formula is obtained; determination of the molecular weight finally leads to the molecular formula. If the alkaloid is optically active, its specific rotation is also measured.

(ii) When an alkaloid contains oxygen, the functional nature of this element is determined:

(*a*) *Hydroxyl group.* The presence of this group may be ascertained by the action of acetic anhydride, acetyl chloride or benzoyl chloride on the alkaloid (acylation must usually be considered in *conjunction with the nature of the nitrogen also present in the* molecule; see iii). When it has been ascertained that hydroxyl groups are present, then their number is also estimated (by acetylation, etc.). The next problem is to decide whether the hydroxyl group is alcoholic or phenolic. It is phenolic if the alkaloid is soluble in sodium hydroxide and reprecipitated by carbon dioxide; also a coloration with ferric chloride will indicate the presence of a phenolic group. If the compound does not behave as a phenol, then the hydroxyl group may be assumed to be alcoholic, and this assumption may be verified by the action of dehydrating agents (most alkaloids containing an alcoholic group are readily dehydrated *by sulphuric acid or phosphorus pentoxide). The behaviour of the* compound towards oxidising agents will also disclose the presence of an alcoholic group.

(*b*) *Carboxyl group.* The solubility of the alkaloid in aqueous sodium carbonate or ammonia indicates the presence of a carboxyl group. The formation of esters also shows the presence of a carboxyl group.

(*c*) *Oxo group.* The presence of an oxo group is readily ascertained by the formation of an oxime, semicarbazone and phenylhydrazone.

(*d*) Hydrolysis of the alkaloid and an examination of the products lead to information that the compound is an ester, lactone, amide, lactam or a betaine.

(*e*) The *Zerewitinoff active hydrogen determination* may be applied to the alkaloid (see Vol. I).

(*f*) *Methoxyl group.* The presence of methoxyl groups and their number may be determined by the *Zeisel method.* The alkaloid is heated with concentrated hydriodic acid at its boiling point (126°); the methoxyl groups are thereby converted into methyl iodide, which is then absorbed by ethanolic silver nitrate and the silver iodide is weighed. Only methoxyl groups have been found in natural alkaloids.

(*g*) *Methylenedioxyl group* (—O·CH_2·O—). The presence of this group is indicated by the formation of formaldehyde when the alkaloid is heated with hydrochloric or sulphuric acid.

(III) *The functional nature of the nitrogen.*

(*a*) The general reactions of the alkaloid with acetic anhydride, methyl iodide and nitrous acid often show the nature of the nitrogen.

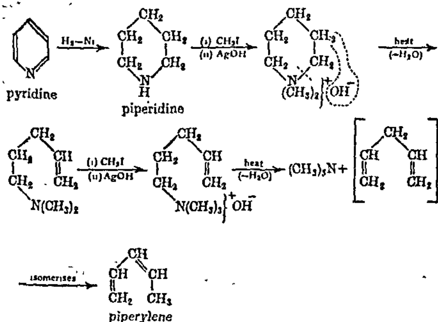
(*b*) Distillation of an alkaloid with aqueous potassium hydroxide usually leads to information regarding the nature and number of alkyl groups attached to nitrogen. The formation (in the volatile products) of methylamine, dimethylamine or trimethylamine indicates respectively the attachment of one, two or three methyl groups to a nitrogen atom; the formation of ammonia shows the presence of an amino group. Only *N*-methyl groups have been shown to be present in alkaloids with one exception, *viz.*, aconitine, which contains an *N*-ethyl group.

(*c*) The presence of *N*-methyl groups and their number may be determined by means of the *Herzig-Meyer method.* When the alkaloid is heated with hydriodic acid at 150–300° under pressure, *N*-methyl groups are converted into methyl iodide (*cf.* the Zeisel method, II*f*).

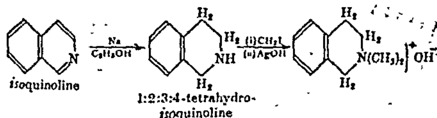
(*d*) The results of hydrolysis will show the presence of an amide, lactam or betaine (*cf.* II*d*).

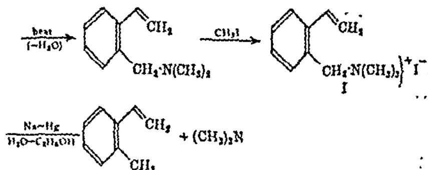
(*e*) *Hofmann's exhaustive methylation method* (1881) is a very important process in alkaloid chemistry, since by its means heterocyclic rings are opened with the elimination of nitrogen, and the nature of the carbon skeleton is thereby obtained. The general procedure is to hydrogenate the heterocyclic ring (if this is unsaturated), then convert this compound to the quaternary methyl-

ammonium hydroxide which is then heated. In this last stage a molecule of water is eliminated, a hydrogen atom *in the β-position* with respect to the nitrogen atom combining with the hydroxyl group, and the ring is opened at the nitrogen atom on the *same* side as the β-hydrogen atom eliminated. The process is then repeated on the product; this results in the complete removal of the nitrogen atom from the molecule, leaving an unsaturated hydrocarbon which, in general, isomerises to a *conjugated* diene (see also Vol. I); *e.g.*,



Hofmann's method fails if there is no β-hydrogen atom available for elimination as water; in such cases the Emde modification (1909, 1912) may be used. In this method the quaternary ammonium halide is reduced with sodium amalgam in aqueous ethanol or catalytically hydrogenated, *e.g.*,

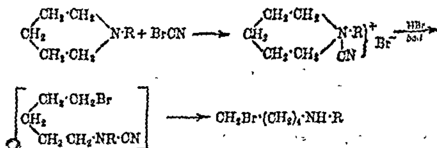




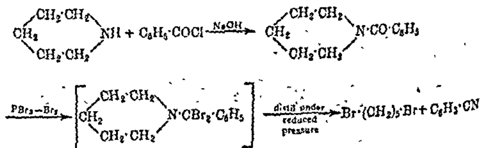
Examination of I shows that β-hydrogen is absent; hence Hofmann's method cannot be used.

Other methods for opening heterocyclic rings containing nitrogen are:

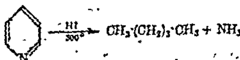
(i) *Von Braun's method* for *tertiary* cyclic amines (see also Vol. I); *e.g.*,



(ii) *Von Braun's method* for *secondary* cyclic amines (see also Vol. I); *e.g.*,



(iii) In a number of cases the ring may be opened by heating with hydriodic acid at 300°, *e.g.*,



(iv) The presence of unsaturation in an alkaloid may be ascertained by the addition of bromine and halogen acids, or by the ability to be hydroxylated with dilute alkaline permanganate. Reduction by means of sodium amalgam, sodium and ethanol, tin and hydrochloric acid, hydriodic acid, etc., also may be used to show the presence of unsaturation. In some cases, reduction may decompose the molecule.

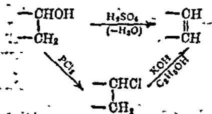
(v) *Oxidation.* This is one of the most valuable means of determining the structure of alkaloids (*cf.* terpenes, §3. VIII). By varying the "strength" of the oxidising agent, it is possible to obtain a variety of products:

(*a*) Mild oxidation is usually effected with hydrogen peroxide, ozone, iodine in ethanolic solution, or alkaline potassium ferricyanide.

(*b*) Moderate oxidation may be carried out by means of acid or alkaline potassium permanganate, or chromium trioxide in acetic acid.

(*c*) Vigorous oxidation is usually effected by potassium dichromate-sulphuric acid, chromium trioxide-sulphuric acid, concentrated nitric acid, or manganese dioxide-sulphuric acid.

This classification is by no means rigid; the "strength" of an oxidising agent depends, to some extent, on the nature of the compound being oxidised. In those cases where it can be done, better results are sometimes achieved by first dehydrating the compound and then oxidising the unsaturated compound thus obtained; oxidation is readily effected at a double bond.



(vi) Fusion of an alkaloid with solid potassium hydroxide often produces relatively simple fragments, the nature of which will give information on the type of nuclei present in the molecule (*cf.* iii*b*).

(vii) *Zinc dust distillation.* This usually gives the same products as (vi), except that when the alkaloid contains oxygen the oxygen is removed.

(viii) *Synthesis.* The foregoing analytical work will ultimately lead to the proposal of a tentative structure (or structures) for the

The molecular formula of ephedrine is $C_{10}H_{15}ON$, and since on oxidation ephedrine forms benzoic acid, the structure therefore contains a benzene ring with only one side-chain. When treated with nitrous acid, ephedrine forms a nitroso-compound; therefore the compound is a secondary amine. Since ephedrine forms a dibenzoyl derivative, one hydroxyl group must be present (one benzoyl group is accounted for by the imino group). Finally, when heated with hydrochloric acid, ephedrine forms methylamine and propiophenone.



$$C_{10}H_{15}ON \xrightarrow{HCl} CH_3{\cdot}NH_2 + C_6H_5{\cdot}CO{\cdot}CH_2{\cdot}CH_3$$

The formation of these products can be explained if the structure of ephedrine is either I or II.

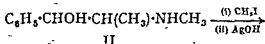




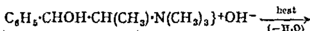


$$\underset{\text{I}}{C_6H_5{\cdot}CHOH{\cdot}CH_2{\cdot}CH_2{\cdot}NH{\cdot}CH_3} \qquad \underset{\text{II}}{C_6H_5{\cdot}CHOH{\cdot}\underset{NH{\cdot}CH_3}{\underset{|}{C}H}{\cdot}CH_3}$$

It has been observed, however, that compounds of structure II undergo the *hydramine fission* to form propiophenone when heated with hydrochloric acid. Thus II is more likely than I. This is supported by the fact when subjected to the Hofmann exhaustive methylation method, ephedrine forms *sym.*-methylphenylethylene oxide, III; this cannot be produced from I, but is to be expected from II.



$$\underset{\text{II}}{C_6H_5{\cdot}CHOH{\cdot}CH(CH_3){\cdot}NHCH_3} \xrightarrow[\text{(ii) AgOH}]{\text{(i) } CH_3I}$$



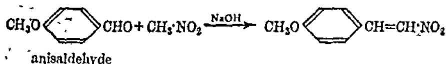
$$C_6H_5{\cdot}CHOH{\cdot}CH(CH_3){\cdot}N(CH_3)_3\}^{+}OH^{-} \xrightarrow[(-H_2O)]{\text{heat}}$$

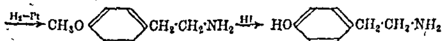


$$\underset{\text{III}}{C_6H_5{\cdot}\overset{\diagup O \diagdown}{CH{\cdot}CH}{\cdot}CH_3} + (CH_3)_3N$$

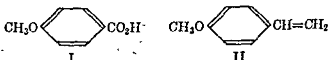
Further support for II is afforded by the following evidence. Structure I contains one asymmetric carbon atom, and so replacement of the hydroxyl group by hydrogen will result in the formation of an optically inactive compound. Structure II, however, contains two asymmetric carbon atoms, and so the replacement of the hydroxyl group by hydrogen should still give a compound that can be optically active. Experimentally it has been found that when this replacement is effected in (—)-ephedrine, the product, deoxyephedrine, is optically active. Thus II agrees with all the known

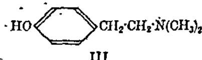
(by the decarboxylation of tyrosine). Tyramine has been synthesised in various ways, *e.g.*,



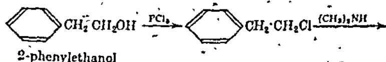


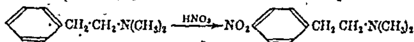
§10. **Hordenine** (β-*p*-hydroxyphenylethyldimethylamine, *Anhaline*), m.p. 117–118°, occurs naturally in germinating barley. The molecular formula of hordenine is $C_{10}H_{15}ON$; the routine tests show that hordenine is a tertiary base and that it contains a phenolic group. Since the methylation of hordenine, followed by oxidation (with alkaline permanganate), gives anisic acid, I, it therefore follows that the hydroxyl group is in the *para*-position with respect to the side-chain. Furthermore, since the methylated compound gives *p*-vinylanisole, II, after the Hofmann exhaustive methylation, the structure of hordenine is probably III.





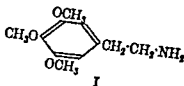
This has been confirmed by synthesis, *e.g.*, Barger (1909):



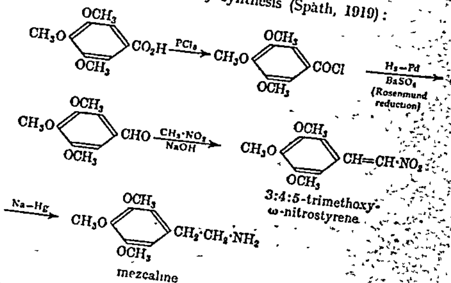




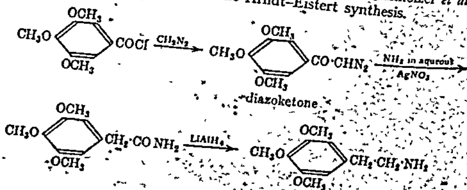
§11. Mezcaline (mescaline), $C_{11}H_{17}O_3N$, b.p. 180–180·5°/12 mm., occurs naturally in "mezcal buttons". The routine tests show that mezcaline contains a primary aliphatic amino-group and three methoxyl groups. On oxidation with alkaline permanganate, mezcaline gives 3:4:5-trimethoxy benzoic acid, and thus the probable structure of mezcaline is I.



This has been confirmed by synthesis (Späth, 1919):



A more recent synthesis of mezcaline is that of Banholzer *et al.* (1952); this makes use of the Arndt–Eistert synthesis.

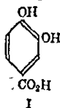


N-Methylmezcaline and *N*-acetylmezcaline also occur naturally in mezcal buttons.

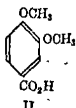
§12. Adrenaline (*Epinephrine*), $C_9H_{13}O_3N$, is a non-steroid hormone. The adrenal medulla is the source of the hormones adrenaline and noradrenaline. Adrenaline was the first hormone to be isolated in a crystalline form (Takamine, 1901; Aldrich, 1901). Adrenaline is active only when given by injection; it raises the blood-pressure, and is used locally to stop hæmorrhage.

Adrenaline is a colourless crystalline solid, m.p. 211°, and dissolves in acids and alkalis (it is insoluble in water); it is also optically active, having a lævorotation.

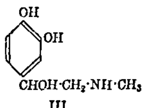
The phenolic character of adrenaline is indicated by its solubility in sodium hydroxide and its reprecipitation by carbon dioxide. Since it gives a green colour with ferric chloride, this led to the suggestion that adrenaline is a catechol derivative. When boiled with aqueous potassium hydroxide, adrenaline evolves methylamine; thus a methylamino group is probably present. On the other hand, when fused with potassium hydroxide, the product is protocatechuic acid, I (Takamine, 1901); methylation, followed by fusion with potassium hydroxide, gives veratric acid, II, and



I

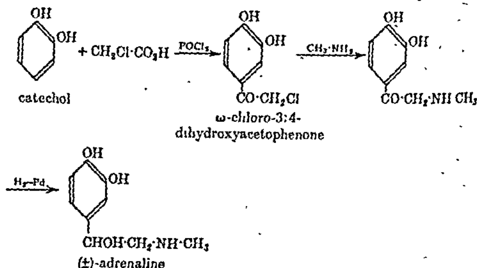


II



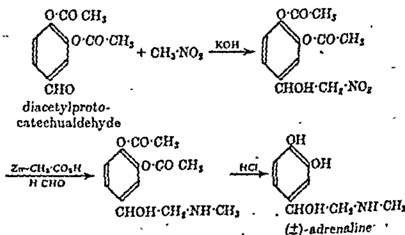
III

trimethylamine (Jowett, 1904). The formation of trimethylamine indicates that the nitrogen atom must occur at the *end* of the side-chain. Since adrenaline is optically active, it must contain at least one asymmetric carbon atom. Now adrenaline contains three hydroxyl groups, two of which are phenolic (as shown by the formation of I and II). The third hydroxyl group was shown to be secondary alcoholic by the fact that when adrenaline is treated with benzenesulphonyl chloride, a tribenzenesulphonyl derivative is obtained which, on oxidation, gives a ketone (Friedmann, 1906). To account for the oxidation of adrenaline to the benzoic acid derivative, the —CHOH— group must be attached directly to the nucleus; had it been $—CH_2{\cdot}CHOH{\cdot}$, then a phenylacetic acid derivative would have been obtained. All the foregoing facts are in keeping with structure III for adrenaline, and this has been confirmed by synthesis by Stolz (1904) and Dakin (1905), with improvements by Ott (1926).

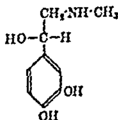


The racemic adrenaline has been resolved by means of (+)-tartaric acid.

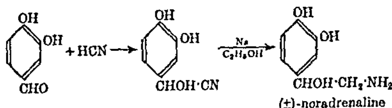
Nagai (1918) has also synthesised adrenaline as follows.



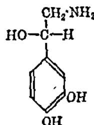
According to Dalgliesh (1953), the configuration of (−)-adrenaline is probably



§12a. **Noradrenaline** (*Norepinephrine*), $C_8H_{11}O_3N$, is also present in the adrenal medulla. The natural compound is lævorotatory, and this (—)-isomer is the most powerful pressor-compound known. The structure of noradrenaline has been established by analytical work similar to that described for adrenaline, and has been confirmed by various syntheses, *e.g.*,

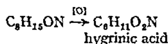


According to Dalgliesh (1953), the configuration of (—)-noradrenaline is



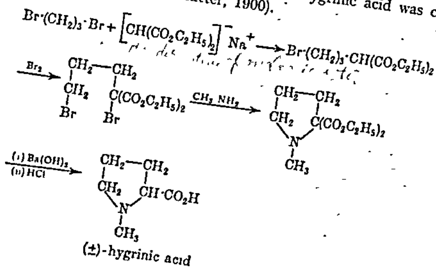
PYRROLIDINE GROUP

§13. **Hygrine**, $C_8H_{15}ON$, b.p. 193–195°, is one of the coca alkaloids. Its reactions show the presence of a keto group and a tertiary nitrogen atom, and when oxidised with chromic acid, hygrinic acid is formed.



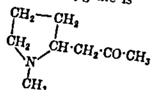
Hygrinic acid was first believed to be a piperidinecarboxylic acid, but comparison with the three piperidine acids showed that this was incorrect. When subjected to dry distillation, hygrinic acid gives *N*-methylpyrrolidine; hence hygrinic acid is an *N*-methylpyrrolidinecarboxylic acid. Furthermore, since the decarboxylation occurs very readily, the carboxyl group was assumed to be in the 2-position (by analogy with the α-amino-acids). This structure,

1-methylpyrrolidine-2-carboxylic acid, for hygrinic acid was confirmed by synthesis (Willstätter, 1900).

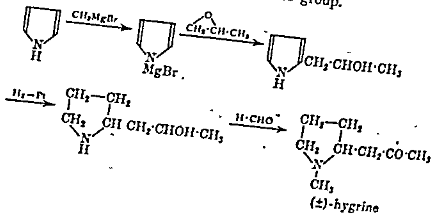


(±)-hygrinic acid

Thus a possible structure for hygrine is

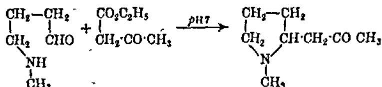


This has been confirmed by synthesis, *e.g.*, Hess (1913); this synthesis starts with pyrrylmagnesium bromide and propylene oxide to form pyrrylpropanol (note the rearrangement that occurs). This compound is then catalytically hydrogenated and then treated with formaldehyde; the imino nitrogen is methylated and the secondary alcoholic is oxidised to a keto group.

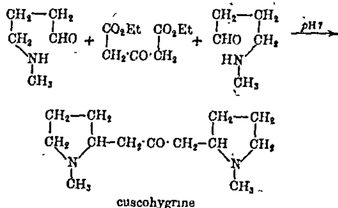


(±)-hygrine

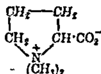
Anet *et al.* (1949) have also synthesised (±)-hygrine by condensing γ-methylaminobutyraldehyde with ethyl acetoacetate in a buffered solution at a *p*H of 7 (physiological conditions).



§13a. Cuscohygrine (*Cuskhygrine*), b.p. 169–170°/23 mm., occurs with hygrine. Its structure is established by the following synthesis (Anet *et al.*, 1949), γ-methylaminobutyraldehyde is condensed with acetonedicarboxylic ester

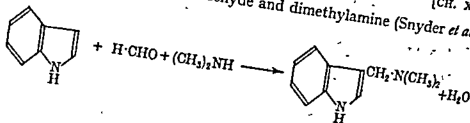


§13b. Stachydrine is obtained from the roots of *Stachys tuberifa*, from orange leaves, etc. It is the betaine (§4. C. XIII) of the quaternary ammonium compound of hygrinic acid.



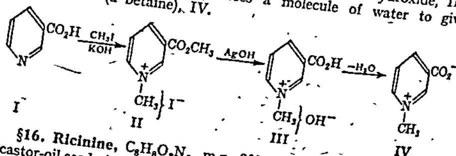
§14. Gramine has been found in barley mutants; it raises the blood-pressure in dogs when administered in small doses. Gramine has been synthesised by allowing indole to stand in an aqueous

solution containing formaldehyde and dimethylamine (Snyder *et al.*, 1944).

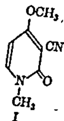


PYRIDINE GROUP

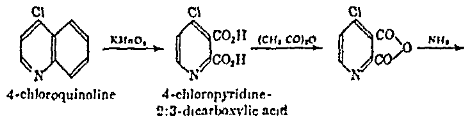
§15. Trigonelline, $C_7H_7O_2N$, m.p. 130°, is widely distributed in plants; the best source is the coffee bean. When boiled with barium hydroxide solution trigonelline produces methylamine; thus the molecule contains an *N*-methylamino group. On the other hand, when heated with hydrochloric acid at 250° under pressure, trigonelline forms methyl chloride and nicotinic acid; this suggests that the alkaloid is the methyl betaine of nicotinic acid. This structure for trigonelline has been confirmed by synthesis (Hantzsch, 1886). When heated with methyl iodide in the presence of potassium hydroxide, nicotinic acid, I, is converted into methyl nicotinate methiodide, II. II, on treatment with "silver hydroxide" solution, forms nicotinic acid methohydroxide, III, which then spontaneously loses a molecule of water to give trigonelline (a betaine), IV.

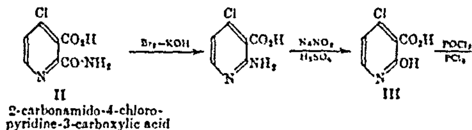


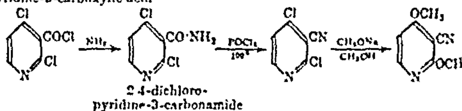
§16. Ricinine, $C_8H_8O_2N_2$, m.p. 201·5°, has been isolated from castor-oil seed; it is not a very toxic alkaloid. Degradative and synthetic work led to the suggestion that I is the structure of ricinine.

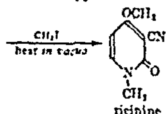


This has been confirmed by synthesis, *e.g.*, Späth *et al.* (1923).

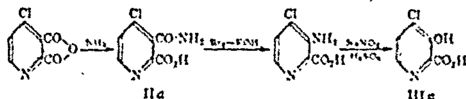




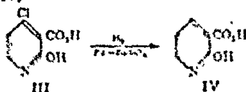




This is not an ... ce II could have been 3-c ... acid, IIa, and conse- qu ...



The structure of III was proved by the fact that on hydrogenation in the presence of Pd—$BaSO_4$, it gave 2-hydroxypyridine-3-carboxylic acid, IV.



A more recent synthesis of ricinine is that of Taylor *et al.* (1956).

$$\text{4-nitro-3-methylpyridine N-oxide} \xrightarrow[H_2SO_4]{K_2Cr_2O_7} \text{4-}NO_2\text{-3-}CO_2H \xrightarrow{CH_3ONa} \text{4-}OCH_3\text{-3-}CO_2H \xrightarrow[HCl]{CH_3OH} \text{4-}OCH_3\text{-3-}CO_2CH_3 \xrightarrow[-33^\circ]{NH_3}$$

$$\text{4-}OCH_3\text{-3-}CONH_2\text{ N-oxide} \xrightarrow[POCl_3]{PCl_5} \text{4-Cl-3-CN-2-Cl pyridine} \xrightarrow{CH_3ONa} \text{4-}OCH_3\text{-3-CN-2-}OCH_3 \xrightarrow{CH_3I} \text{4-}OCH_3\text{-3-CN-1-}CH_3\text{-2-pyridone}$$

§17. Areca (or Betel) nut alkaloids. The betel nut is the source of a number of alkaloids which are all partially hydrogenated derivatives of nicotinic acid, *e.g.*,

guvacine, m.p. 271-272°; guvacoline, b.p. 114°/13 mm.; arecaidine, m.p. 223°; arecoline, b.p. 209°

Let us consider arecaidine; its molecular formula is $C_7H_{11}O_2N$. When distilled with zinc dust, guvacine gives 3-methylpyridine; therefore this alkaloid is a pyridine derivative. Now guvacine is converted into arecaidine on heating with potassium methyl sulphate and sodium methoxide (Jahns, 1888, 1890); thus arecaidine is a methyl derivative of guvacine, and consequently is also a pyridine derivative. The usual tests show that arecaidine contains one carboxyl group, an *N*-methyl group, and one double bond; hence the formula for arecaidine may be written as $C_5H_7N(CH_3)\cdot CO_2H$. Since the alkaloid is a pyridine derivative, the fragment C_5H_7N could be tetrahydropyridine. This was proved to be so by synthesis, and at the same time the positions of the double bond and carboxyl group were also established (Wohl *et al.*, 1907). Acraldehyde, I, on treatment with ethanol in the presence of hydrogen chloride,

$CHO{-}CH{=}CH_2 + 2C_2H_5OH + HCl \longrightarrow CH(OC_2H_5)_2{-}CH_2{-}CH_2Cl \xrightarrow{CH_3NH_2} [(C_2H_5O)_2CH{-}CH_2{-}CH_2]_2N{-}CH_3$

I II III

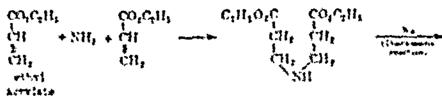
$\xrightarrow{HCl}$ $[(CHO{-}CH_2{-}CH_2)_2N{-}CH_3] \longrightarrow$ IV $\xrightarrow[(ii)\ SOCl_2]{(i)\ NH_2OH}$ V $\xrightarrow{HCl}$

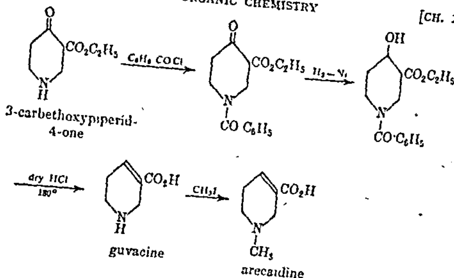
IV V

VI *or* VI *a*

forms 3-chloropropionaldehyde acetal, II. II reacts with methylamine to form β-methyliminodipropionaldehyde tetra-acetal, III, which, on treatment with concentrated hydrochloric acid, ring closes to form 1:2:5:6-tetrahydro-1-methylpyridine-3-aldehyde, IV. This gives the cyano compound V on treatment with hydroxylamine, followed by dehydration of the oxime with thionyl chloride, and V is then converted into arecaidine by hydrolysis. Arecaidine is VI, or possibly VIa, the dipolar ion structure (*cf.* amino-acids and betaines).

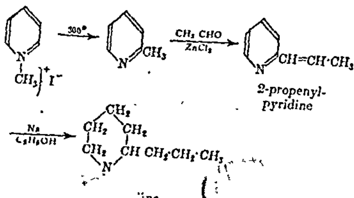
A more recent synthesis of arecaidine (and guvacine) is that of McElvain *et al.* (1946).





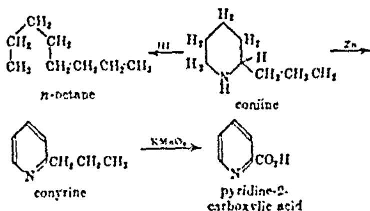
§18. Hemlock alkaloids. The most important alkaloid of this group is **coniine**; it was the first alkaloid to be synthesised. Oil of hemlock was drunk by Socrates when he was condemned to death in 399 B.C.

(+)-Coniine, $C_8H_{17}N$, b.p. 166–167°, is the form that occurs in oil of hemlock. When distilled with zinc dust, coniine is converted into conyrine, $C_8H_{11}N$ (Hofmann, 1884). Since the oxidation of conyrine with permanganate gives pyridine-2-carboxylic acid (α-picolinic acid), it follows that a pyridine nucleus is present with a side-chain in the 2-position. Thus coniine is probably a piperidine derivative with a side-chain in the 2-position. This side-chain must contain three carbon atoms, since two are lost when conyrine is oxidised. This side-chain is therefore either *n*-propyl or *iso*-propyl, and it was actually shown to be *n*-propyl by the fact that

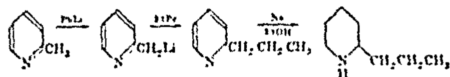


when heated with hydriodic acid at 300° under pressure, coniine forms *n*-octane. Had the side-chain been *iso*propyl, then the expected product would be *iso*-octane. From this evidence it therefore follows that coniine is 2-*n*-propylpiperidine, and this has been confirmed by synthesis (Ladenburg, 1885). The racemic coniine was resolved by means of (+)-tartaric acid, and the (+)-coniine so obtained was found to be identical with the natural compound.

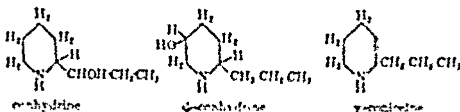
The reactions of coniine described above can therefore be formulated as follows:



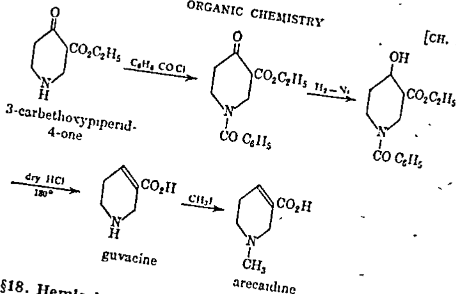
Coniine has also been synthesised from 2-methylpyridine and phenyl-lithium as follows (Bergmann *et al.*, 1932):



Other hemlock alkaloids are:

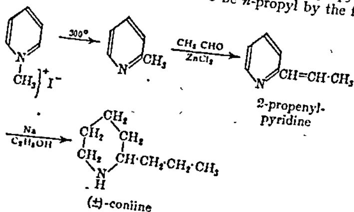


§19. Pomegranate alkaloids. The root bark of the pomegranate tree contains a number of alkaloids, the most important of



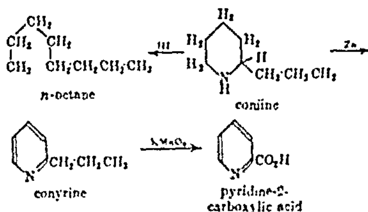
§18. Hemlock alkaloids. The most important alkaloid of this group is **coniine**; it was the first alkaloid to be synthesised. Oil of hemlock was drunk by Socrates when he was condemned to death in 399 B.C.

(+)-**Coniine,** $C_8H_{17}N$, b.p. 166–167°, is the form that occurs in oil of hemlock. When distilled with zinc dust, coniine is converted into conyrine, $C_8H_{11}N$ (Hofmann, 1884). Since the oxidation of conyrine with permanganate gives pyridine-2-carboxylic acid (α-picolinic acid), it follows that a pyridine nucleus is present with a side-chain in the 2-position. Thus coniine is probably a piperidine derivative with a side-chain in the 2-position. This side-chain must contain three carbon atoms, since two are lost when conyrine is oxidised. This side-chain is therefore either *n*-propyl or *iso*-propyl, and it was actually shown to be *n*-propyl by the fact that

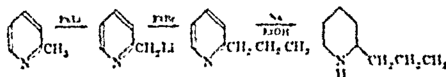


when heated with hydriodic acid at 300° under pressure, coniine forms *n*-octane. Had the side-chain been *iso*propyl, then the expected product would be *iso*-octane. From this evidence it therefore follows that coniine is 2-*n*-propylpiperidine, and this has been confirmed by synthesis (Ladenburg, 1885). The racemic coniine was resolved by means of (+)-tartaric acid, and the (+)-coniine so obtained was found to be identical with the natural compound.

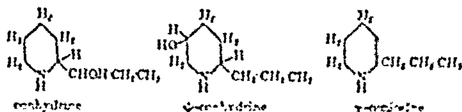
The reactions of coniine described above can therefore be formulated as follows:



Coniine has also been synthesised from 2-methylpyridine and phenyl-lithium as follows (Bergmann *et al.*, 1932).



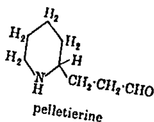
Other hemlock alkaloids are:

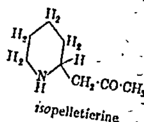


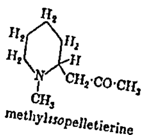
conhydrine ψ-conhydrine γ-coniceine

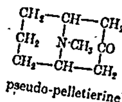
§19. Pomegranate alkaloids. The root bark of the pomegranate tree contains a number of alkaloids, the most important of

which is pelletierine; three others are *iso*pelletierine, methyl*iso*pelletierine and pseudo-pelletierine. The last of these is related to atropine (§22).

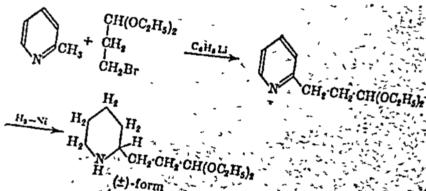




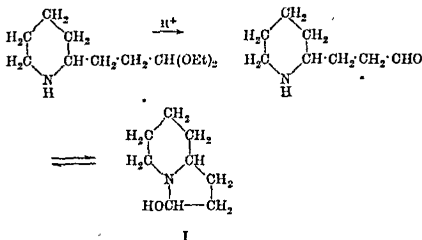




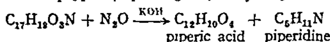
Pelletierine acetal has been synthesised by Spielman *et al.* (1941) by the action of 3-bromopropionaldehyde acetal on 2-methylpyridine (α-picoline) in the presence of phenyl-lithium, followed by catalytic reduction.



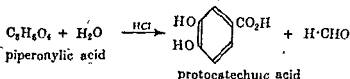
elletierine acetal was also prepared by Wibaut *et al.* (1940) who tempted to hydrolyse it to the free aldehyde; they obtained ly viscous oils. Spielman *et al.* also failed to obtain the free lehyde. Beets (1943) has therefore suggested that pelletierine , and probably does, exist as some bicyclic structure such as I.



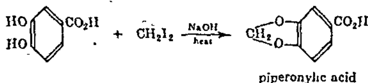
§20. Piperine, $C_{17}H_{19}O_3N$, m.p. 128–129·5°, occurs in pepper, especially black pepper (*Piper nigrum*). Hydrolysis of piperine



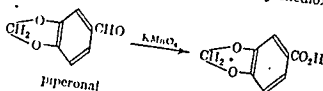
with alkali gives piperic acid and piperidine; thus the alkaloid is the piperidine amide of piperic acid (Babo *et al.*, 1857). Since piperidine is hexahydropyridine, the structure of piperine rests on the elucidation of that of piperic acid. The routine tests show that piperic acid contains one carboxyl group and two double bonds. When oxidised with permanganate, piperic acid gives first piperonal and then piperonylic acid. The structure of the latter is deduced from the fact that when heated with hydrochloric acid at 200° under pressure, piperonylic acid forms protocatechuic acid (3,4-dihydroxybenzoic acid) and formaldehyde.



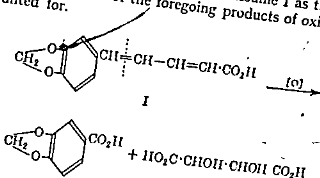
Since one atom of carbon is eliminated, and there are no free hydroxyl groups in piperonylic acid, the structure of this acid is probably the methylene ether of protocatechuic acid, *i.e.*, piperonylic acid is 3:4-methylenedioxybenzoic acid; this has been confirmed by synthesis:



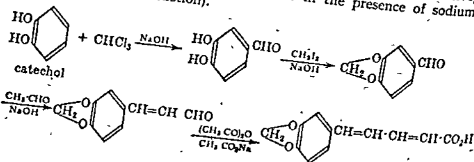
Furthermore, since piperonal (an aldehyde) gives piperonylic acid on oxidation, piperonal is therefore 3,4-methylenedioxybenzaldehyde.



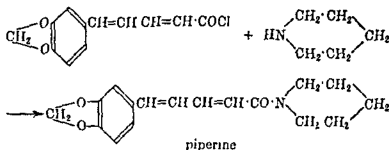
From these results of oxidative degradation, it therefore follows that piperic acid is a benzene derivative containing only one side-chain. It is this side-chain that contains the two double bonds (the ready addition of four bromine atoms shows the presence of two *ethylenic* bonds), and since the careful oxidation of piperic acid gives tartaric acid in addition to piperonal and piperonylic acid, the side-chain is a "straight" chain. If we assume I as the structure of piperic acid, then all of the foregoing products of oxidation may be accounted for.



This has been confirmed by synthesis (Ladenburg *et al.*, 1894); piperonal (prepared *via* the Reimer–Tiemann reaction) is condensed with acetaldehyde in the presence of sodium hydroxide (Claisen–Schmidt reaction), and the product (a cinnamaldehyde derivative) is then heated with acetic anhydride in the presence of sodium acetate (Perkin reaction).



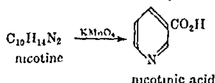
When the acid chloride of piperic acid (prepared by the action of phosphorus pentachloride on the acid) is heated with piperidine in benzene solution, piperine is formed; thus piperine is the piperidine amide of piperic acid.



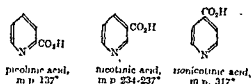
PYRROLIDINE-PYRIDINE GROUP

§21. Tobacco alkaloids. Many alkaloids have been isolated from the tobacco leaf, *e.g.*, nicotine, nicotimine (anabasine), nornicotine, etc.

Nicotine, $C_{10}H_{14}N_2$, b.p. 247°, is the best known and most widely distributed of the tobacco alkaloids; it occurs naturally as the (−)-form. When oxidised with dichromate-sulphuric acid (or permanganate or nitric acid), nicotine forms nicotinic acid (Huber, 1867).

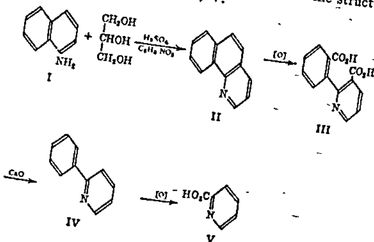


It is instructive, at this point, to see how the orientations of the three isomeric pyridinecarboxylic acids have been elucidated

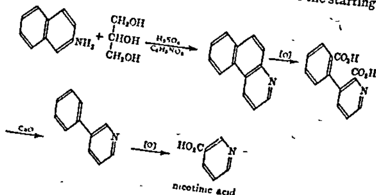


Picolinic acid 1-Naphthylamine, I, when subjected to the Skraup synthesis (see Vol. I), is converted into 7:8-benzoquinoline, II (this

structure is established by its synthesis). II, on vigorous oxidation with alkaline permanganate, gives the dicarboxylic acid III which, when decarboxylated by heating with calcium oxide, is converted into 2-phenylpyridine, IV This, on further oxidation with permanganate, gives a pyridinecarboxylic acid which must, from the structure of IV, be the 2-acid, *i e.*, picolinic acid, V.



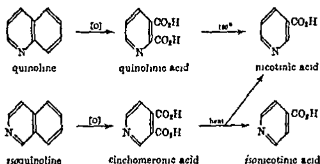
Nicotinic acid This has been shown to be pyridine-3-carboxylic acid by a similar set of reactions, except that in this case the starting material is 2-naphthylamine.



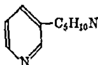
isoNicotinic acid. This third isomer is therefore pyridine-4-carboxylic acid.

An alternative proof for the orientations of these three acids is based on the structures of quinoline and *iso*quinoline (which have been established by synthesis). Oxidation of quinoline with alkaline permanganate gives quinolinic acid which, by its method of preparation, must be pyridine-2 3-dicarboxylic acid. When quinolinic acid is heated to 190°, one carboxyl group is lost to produce nicotinic acid; thus nicotinic acid must be either pyridine-2- or -3-carboxylic acid. *iso*Quinoline, on oxidation with alkaline permanganate, produces

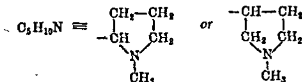
cinchomeronic acid, which must therefore be pyridine-3·4-dicarboxylic acid. This, on gentle heating, gives a mixture of nicotinic and *iso*-nicotinic acids; thus nicotinic acid must be the 3-acid, and *iso*nicotinic acid the 4-acid. Hence picolinic acid is pyridine-2-carboxylic acid.



Returning to the structure of nicotine, since nicotinic acid is a product of oxidation, the alkaloid therefore contains a pyridine nucleus with a complex side-chain in the 3-position. Thus we may write the formula of nicotine as

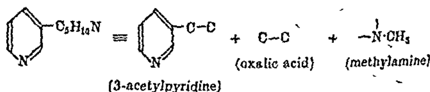


Because of its formula, this side-chain was originally believed to be piperidine, but further work showed that this was incorrect. When nicotine zinc chloride is distilled, the products are pyridine, pyrrole and methylamine (Laiblin, 1879). This suggests that the side-chain $C_5H_{10}N$ is a pyrrole derivative. Furthermore, when nicotine is heated with concentrated hydriodic acid at 150° (Herzig–Meyer method), methyl iodide is formed. Thus the side-chain contains an *N*-methyl group. It therefore appears that the side-chain could be *N*-methylpyrrolidine, but its point of attachment to the pyridine nucleus could be either 2 or 3 on the evidence obtained so far:

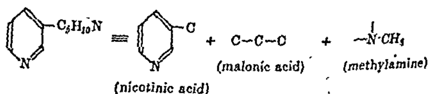


The correct structure of nicotine was obtained by Pinner (1892, 1893). Treatment of nicotine with bromine in acetic

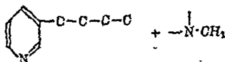
acid gives, among other products, the hydrobromide perbromide, $C_{10}H_{10}ON_2Br_2.HBr.Br_2$, which, when treated with aqueous sulphurous acid, is converted into dibromocotinine, $C_{10}H_{10}ON_2Br_2$. This, on heating with a mixture of sulphurous and sulphuric acids at 130–140°, forms 3-acetylpyridine, oxalic acid and methylamine. Thus the structure of nicotine must account for the following skeleton structures:



Now bromine, in the presence of hydrobromic acid, converts nicotine into dibromoticonine, $C_{10}H_8O_2N_2Br_2$, which, on heating with barium hydroxide solution at 100°, forms nicotinic acid, malonic acid and methylamine. Hence the structure of nicotine must also account for the following skeleton structures:

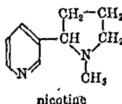


These two sets of reactions, taken in conjunction with one another, are satisfied by the following skeleton for nicotine:



The problem now is: Where is the position of the *N*-methyl group? Nicotine behaves as a *di-tertiary base*, and forms two isomeric "methyl iodide addition products". Thus the nitrogen atom in the side-chain must be of the type —C—N(CH_3)—C—. Furthermore, it is extremely difficult to reduce nicotine beyond hexahydronicotine (the pyridine part is reduced to piperidine). Hence the side-chain must be saturated, and this can only be so if the side-chain is cyclic, *i.e.*, *N*-methylpyrrolidine ($C_5H_{11}N \equiv C_4H_8{\cdot}NCH_3 \equiv C_4H_8$). The presence of this pyrrolidine nucleus also accounts

for the formation of pyrrole when nicotine zincichloride is distilled (see above). All the foregoing facts are satisfied by the following structure for nicotine.



nicotine

On this basis, Pinner's work may be formulated:

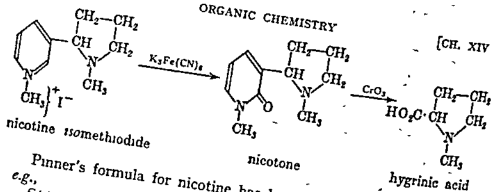
(i) nicotine $\xrightarrow[\text{(ii) } H_2SO_3]{\text{(i) } Br_2—CH_3·CO_2H}$ dibromocotinine $\xrightarrow[H_2SO_4]{H_2SO_3}$

3-acetylpyridine + CO_2H—CO_2H + $CH_3·NH_2$

(ii) nicotine $\xrightarrow[HBr]{Br_2}$ dibromoticonine $\xrightarrow{Ba(OH)_2}$

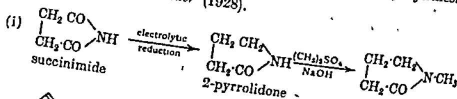
nicotinic acid (CO_2H) + $HO_2C·CH_2·CO_2H$ + $CH_3·NH_2$

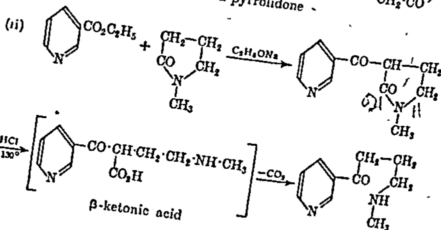
The most direct analytical evidence for the presence of the pyrrolidine nucleus has been given by Karrer (1925, 1926); nicotine hydriodide forms nicotine *iso*methiodide when warmed with methyl iodide and this, on oxidation with potassium ferricyanide, is converted into nicotone which, on oxidation with chromium trioxide, gives hygrinic acid (§13).

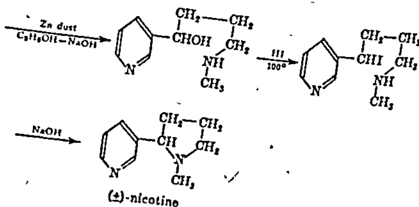


Pinner's formula for nicotine has been confirmed by synthesis, e.g.,

Späth and Bretschneider (1928).







This was resolved by means of (+)-tartaric acid; the synthetic (−)-nicotine is identical with the natural compound.

Craig (1933).

nicotinonitrile + $BrMgCH_2{\cdot}CH_2{\cdot}CH_2OC_2H_5$ (γ-ethoxypropylmagnesium bromide) ⟶ 3-pyridyl γ-ethoxypropyl ketone $\xrightarrow{NH_2OH}$ oxime $\xrightarrow[CH_3{\cdot}CO_2H]{Zn}$ amine $\xrightarrow[150-155°]{HBr}$ [bromo-amine] ⟶ (±)-nornicotine $\xrightarrow[(ii)\ NaOH]{(i)\ CH_3I}$ (±)-nicotine

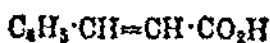
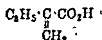
§22. **Solanaceous alkaloids.** This group includes atropine, hyoscyamine and scopolamine (hyoscine).

Atropine, $C_{17}H_{23}O_3N$, m.p. 118°, occurs in deadly nightshade (*Atropa belladonna*) together with hyoscyamine. Hyoscyamine is optically active (lævorotatory), but readily racemises to atropine when warmed in an ethanolic alkaline solution; thus atropine is (±)-hyoscyamine.

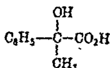
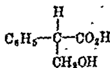
When warmed with barium hydroxide solution, atropine is hydrolysed to (±)-tropic acid and tropine (an alcohol); thus atropine is the tropine ester of tropic acid.

(±)-Tropic acid, $C_9H_{10}O_3$, m.p. 117°, is a saturated compound (it does not add on bromine); the usual tests show that it contains one carboxyl group and one alcoholic group. When heated strongly, tropic acid loses a molecule of water to form atropic acid, $C_9H_8O_2$,

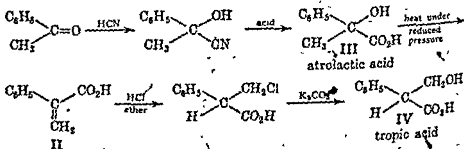
and this, on oxidation, gives benzoic acid. Thus tropic and atropic acids contain a benzene ring with one side-chain. It therefore

follows that atropic acid could be either I or II. Since, however, I is known to be cinnamic acid, II must be atropic acid. Addition of a molecule of water to II would therefore give tropic acid which,

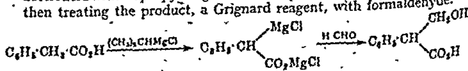
consequently, must be either III or IV. Tropic acid has been shown to be IV by synthesis, *e.g.*, Mackenzie and Wood (1919), starting from acetophenone.



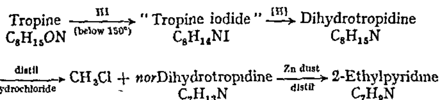
III is atrolactic acid, and its dehydration to II confirms the structure of atropic acid. It should also be noted that the addition of hydrogen chloride takes place contrary to Markownikoff's rule (see unsaturated acids, Vol. I); had the addition been in accordance with the rule, then atrolactic acid would have again been obtained. [illegible] carbon atom which [illegible]

quinine.

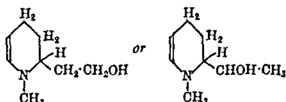
Blicke *et al.* (1952) have synthesised tropic acid by boiling phenylacetic acid with *iso*propylmagnesium chloride in ethereal solution, and then treating the product, a Grignard reagent, with formaldehyde.



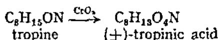
Tropine (tropanol), $C_8H_{15}ON$, m.p. 63°, behaves as a saturated compound which contains an alcoholic group. The structure of tropine was investigated by Ladenburg (1883, 1887), who showed that the molecule contained a reduced pyridine nucleus:



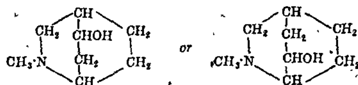
"Tropine iodide" is formed by the replacement of the alcoholic group in tropine by an iodine atom, which is then replaced by hydrogen to form dihydrotropidine (tropane). The formation of methyl chloride indicates the presence of an *N*-methyl group, and the isolation of 2-ethylpyridine shows the presence of this nucleus (in a reduced form). Largely on this evidence, Ladenburg was led to suggest the following alternative formulæ for tropine:



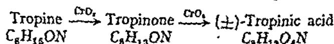
Merling (1891), by the oxidation of tropine with chromium trioxide, obtained (±)-tropinic acid.



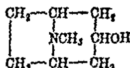
Tropinic acid is a dicarboxylic acid, and since there is no loss of carbon in its formation, the hydroxyl group in tropine must therefore be in a ring system. Thus Ladenburg's formula is untenable, and so Merling proposed the following structures for tropine:



Willstatter (1895–1901) then examined the oxidation products of tropine obtained as follows:



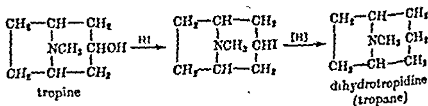
Tropinone behaved as a ketone; thus tropine is a secondary alcohol (*cf.* Merling's formula). Willstätter (1897) also showed that tropinone forms a dibenzylidene derivative with benzaldehyde, and a di-oximino derivative when treated with amyl nitrite and hydrochloric acid. Thus tropinone contains the $\cdot CH_2 \cdot CO \cdot CH_2 \cdot$ grouping, and so it follows that Merling's formula is also untenable. Willstätter therefore proposed three possible structures for tropine, but eliminated two by the consideration of various reactions of tropine, and was left with the following (which contains a pyridine and a pyrrole nucleus with the nitrogen atom common to both):

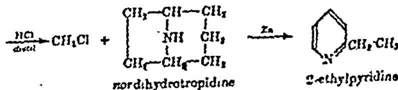


Not only did this fit the facts best, but it was also supported by the following evidence: (i) Exhaustive methylation of tropine gives tropilidene (*cyclo*heptatriene), C_7H_8. (ii) Exhaustive methylation of tropinic acid gives an unsaturated dicarboxylic acid which, on reduction, forms pimelic acid.

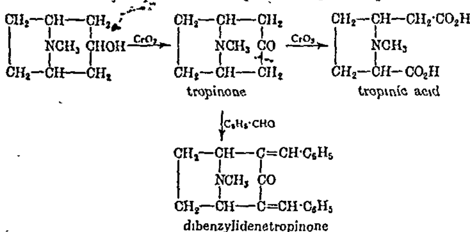
All the foregoing reactions of tropine can be readily explained on the Willstätter formula.

Formation of 2-ethylpyridine from tropine.

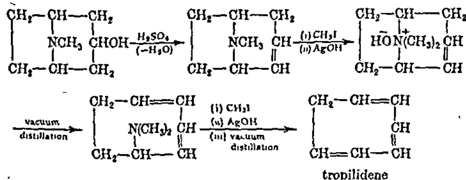




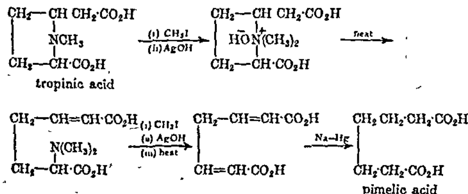
Formation of tropinone and tropinic acid from tropine.



Formation of tropilidene from tropine.

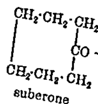


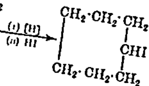
Formation of pimelic acid from tropinic acid.

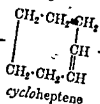


The structure of tropine has been confirmed by synthesis, one by Willstätter (1900–1903), and the other by Robinson (1917).

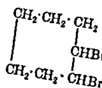
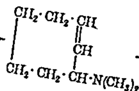
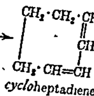
Willstätter's synthesis.



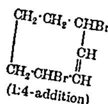


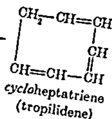





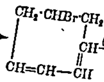
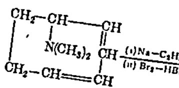




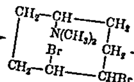
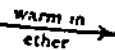


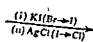


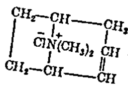
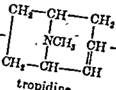


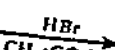


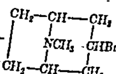
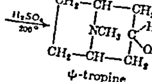



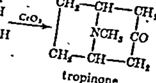



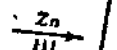


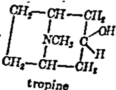





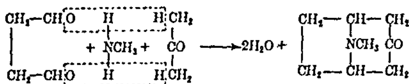




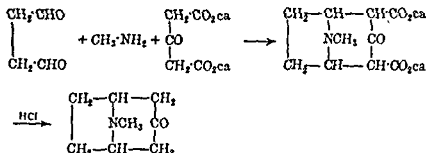


Robinson's synthesis.

When a mixture of succinaldehyde, methylamine and acetone is allowed to stand in water for thirty minutes, tropinone is produced in very small yield.

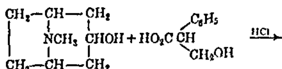


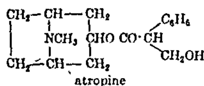
A much better yield (40 per cent.) is obtained by using calcium acetonedicarboxylate or ethyl acetonedicarboxylate instead of acetone; the calcium salt or ester so produced is converted into tropinone by warming with hydrochloric acid, *e.g.* (ca = Ca/2):



Schöpf *et al.* (1935) have obtained a yield of 70-85 per cent. by carrying out Robinson's synthesis at a *p*H of 7 (see also §28).

The final problem is to combine tropine with tropic acid, this has been done by heating the two together in the presence of hydrogen chloride (Fischer-Speier esterification; see Vol. I).





Stereochemistry of the tropines. Tropinone can be reduced to tropine, together with a small amount of ψ-tropine, by means of a

metal and acid, the best combination being zinc dust and hydriodic acid, or by means of electrolytic reduction. On the other hand, reduction with sodium amalgam converts tropinone into ψ-tropine. According to Mirza (1952), lithium aluminium hydride reduces tropinone quantitatively to ψ-tropine, but according to Beckett *et al.* (1957), 54 per cent. of ψ-tropine and 45 per cent. of tropine are obtained. A larger yield of the former (69 per cent.) is obtained with sodium borohydride, and reduction with sodium and *iso*butanol (in toluene) gives the maximum yield of ψ-tropine (88 per cent.).

Tropine and ψ-tropine are geometrical isomers, one isomer having the hydrogen atom on C_3 on the same side as the nitrogen bridge, and the other isomer has this hydrogen atom on the opposite side (*cf.* the borneols, §23b. VIII); Fig. 1 shows the two possible forms.

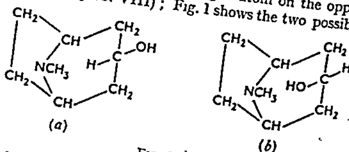


Fig. 14.1.

Neither of these forms is optically active, since the molecule has a plane of symmetry. C_1 and C_5 are asymmetric, but the molecule is optically inactive by internal compensation (see §7b. II), and so each isomer is a *meso*-form; C_3 is pseudo-asymmetric (see §8. IV). It should also be noted that another pair of *optically active forms* would exist if the fusion of the nitrogen bridge were *trans*; this, however, is not possible (*cf.* camphor, §23a. VIII; also cocaine, §23).

The problem now is to decide which geometrical isomer (of the two forms shown in Fig. 1) is tropine and which is ψ-tropine. Fodor (1953) has given evidence to show that ψ-tropine is the *syn*-compound (nitrogen bridge and hydroxyl group are in the *cis*-position; Fig. 1b), and that tropine is the *anti*-compound (nitrogen bridge and hydroxyl group are in the *trans*-position; Fig. 1a).

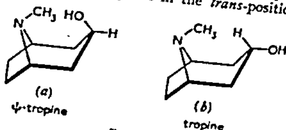


Fig. 14.2.

The problem, however, is more involved than this, since the conformation of the piperidine ring has also to be considered. Fodor gives the configuration of the piperidine ring as the boat form in both isomers (Fig. 2).

Zenitz *et al.* (1952) and Clemo *et al.* (1953) support these configurations from evidence obtained by measurements of the dipole moments of these two isomers; ψ-tropine has been shown to have a higher dipole moment than tropine. Zenitz *et al.* have also shown from infra-red absorption spectra measurements that ψ-tropine has intramolecular hydrogen bonding; this is only possible in the *syn*-form. Bose *et al.* (1953), however, have assumed the chair form for the piperidine ring by analogy with the chair conformation of *cyclo*hexane compounds and pyranosides (see §11. IV). Thus these authors have suggested that ψ-tropine is Fig. 3 (*a*), in which the hydroxyl group is equatorial, and that tropine is Fig. 3 (*b*), in which the hydroxyl group is axial.

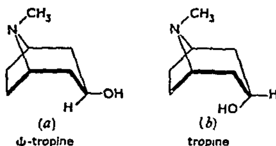
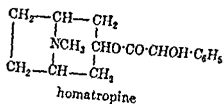


Fig. 14.3

If these be the configurations, then it is difficult to explain Fodor's work (which involves rearrangements), and also the fact that there is intramolecular hydrogen bonding in ψ-tropine. Sparke (1953) has suggested that the chair form can readily change into the boat form; this would then explain the intramolecular hydrogen bonding. Archer and Lewis (1954) also adopt this explanation, but make the assumption that the bond energy involved in the hydrogen bond is sufficient to transform, at least partially, the more stable chair form into the less stable boat form; in ψ-tropine the chair and boat forms are in mobile equilibrium, the latter being the predominant form.

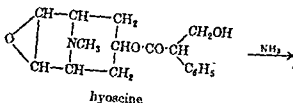
§22a. Tropeines and pseudotropeines. These are synthetic esters formed respectively from tropine and ψ-tropine with various organic acids. The tropeines (including atropine itself) are powerful mydriatics (pupil dilators) and feeble anæsthetics; the ψ-tropeines are the reverse. One of the most important tropeines is *homatropine*

(*mandelyltropeine*), which is prepared by combining tropine with mandelic acid.

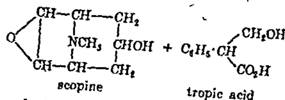


homatropine

§22b. **Hyoscine (scopolamine)**, $C_{17}H_{21}O_4N$, is a syrup and is lævorotatory; it is obtained from various sources, *e.g.*, *Datura Metel*. Hyoscine is a constituent of travel sickness tablets, and when administered with morphine, produces "twilight sleep". Hyoscine is the (−)-tropic ester of the aminoalcohol *scopine*; these two compounds are produced by the hydrolysis of hyoscine with ammonia.

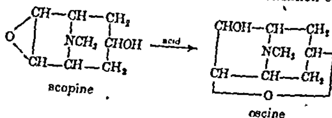


hyoscine



scopine tropic acid

More vigorous hydrolysis of hyoscine with acids or alkalis produces *oscine* (*scopoline*), which is formed by the isomerisation of scopine.

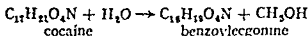


scopine oscine

It is interesting to note, in this connection, that the action of *ethanolic* sodium hydroxide on (−)-hyoscine at room temperature causes the latter to racemise to (±)-hyoscine.

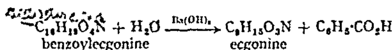
§23. **Coca alkaloids.** In this group occur cocaine, benzoylecgonine, tropacocaine, hygrine (§13), cuscohygrine (§13a), etc.

(−)-Cocaine, $C_{17}H_{21}O_4N$, m.p. 98°, occurs in coca leaves; it is sparingly soluble in water, but its hydrochloride is quite soluble and is used as a local anæsthetic. When heated with water, cocaine is hydrolysed to methanol and benzoylecgonine.



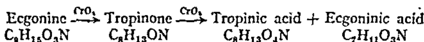
$$\underset{\text{cocaine}}{C_{17}H_{21}O_4N} + H_2O \rightarrow \underset{\text{benzoylecgonine}}{C_{16}H_{19}O_4N} + CH_3OH$$

Thus cocaine contains a carbomethoxyl group, and benzoylecgonine a carboxyl group. When benzoylecgonine is heated with barium hydroxide solution, further hydrolysis occurs, the products obtained being benzoic acid and ecgonine.



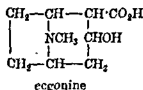
$$\underset{\text{benzoylecgonine}}{C_{16}H_{19}O_4N} + H_2O \xrightarrow{Ba(OH)_2} \underset{\text{ecgonine}}{C_9H_{15}O_3N} + C_6H_5{\cdot}CO_2H$$

Ecgonine shows the reactions of an alcohol, and so benzoylecgonine is the benzoyl derivative of a hydroxycarboxylic acid. The structure of ecgonine has been deduced from the nature of the products obtained by oxidation, *viz.*,



$$\underset{C_9H_{15}O_3N}{\text{Ecgonine}} \xrightarrow{CrO_3} \underset{C_8H_{13}ON}{\text{Tropinone}} \xrightarrow{CrO_3} \underset{C_8H_{13}O_4N}{\text{Tropinic acid}} + \underset{C_7H_{11}O_3N}{\text{Ecgoninic acid}}$$

From these results, it follows that ecgonine contains the tropane structure and that the alcoholic group must be in the same position as in tropine (§22). Now in the formation of tropinone from ecgonine, a carboxyl group is lost (as we have seen, ecgonine contains a carboxyl group). Thus the carboxyl group is in a position such that the oxidation of the secondary alcoholic group in ecgonine to a keto group is accompanied by the elimination of the carboxyl group. This type of elimination is characteristic of β-ketonic acids, and this interpretation of the results is confirmed by the fact that Willstätter *et al.* (1898) actually observed the formation of an unstable β-ketonic acid which lost carbon dioxide to give tropinone. Thus ecgonine is:



ecgonine

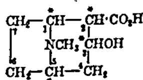
On this basis, the foregoing reactions may therefore be written:

```
CH2—CH—CH·CO2CH3                              CH2—CH—CH·CO2H
|    |    |                                   |    |    |
|   NCH3 CHO·CO C6H5  --H2O-->  CH3OH  +      |   NCH3 CHO·CO C6H5
|    |    |                                   |    |    |
CH2—CH—CH2                                    CH2—CH—CH2
   cocaine                                       benzoylecgonine

           CH2—CH—CH·CO2H                              [CH2—CH—CH·CO2H]
           |    |    |                                 [|    |    |    ]
--Ba(OH)2-->|   NCH3 CHOH  +  C6H5·CO2H  --CrO3-->      [|   NCH3 CO    ]
           |    |    |                                 [|    |    |    ]
           CH2—CH—CH2                                  [CH2—CH—CH2     ]
              ecgonine

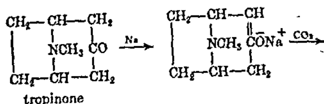
        CH2—CH—CH2              CH2—CH·CH2·CO2H        CH2—CH·CH2·CO2H
        |    |    |             |                      |
--CO2-->|   NCH3 CO  --CrO3-->  |   NCH3          +    |   NCH3
        |    |    |             |                      |
        CH2—CH—CH2              CH2—CH·CO2H            CH2—CO
          tropinone               tropinic acid          ecgoninic acid
```

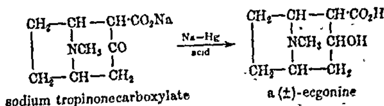
The structure of ecgonine has been confirmed by synthesis (Willstätter *et al.*, 1901); the starting point is tropinone (see §22 for its synthesis). Before describing this synthesis, let us first examine the structure of ecgonine from the stereochemical point



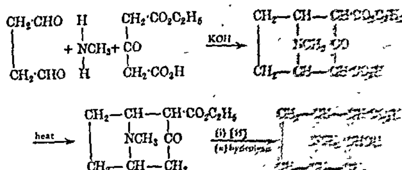
of view; it will be seen that there are four dissimilar asymmetric carbon atoms present (*), and so there are $2^4 = 16$ optically active forms (eight pairs of enantiomorphs) possible (*cf.* tropine, §22). Since, however, only the *cis* fusion of the nitrogen bridge is possible in practice, C_1 and C_5 therefore have only one configuration (the *cis*-form), and so there are only eight optically active forms (four pairs of enantiomorphs) actually possible (*cf.* camphor, §23a, VIII); three pairs of enantiomorphs have been prepared synthetically.

In the original synthesis of Willstätter, the racemic ecgonine obtained was not identical with the (—)-ecgonine from (—)-cocaine, but its chemical properties were the same.





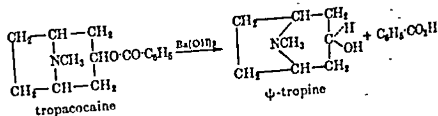
Later, Willstätter *et al.* (1921) synthesised ecgonine by means of the Robinson method (see §22):



The final product was shown to be a mixture of [illegible], (±)-ecgonine, (±)-ψ-ecgonine, and a [illegible] (Willstätter *et al.*, 1923). The racemic [illegible] was resolved, and the (−)-form esterified with [illegible]; the product was (−)-cocaine.

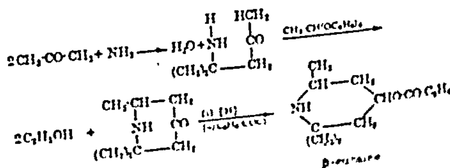
connection is that Einhorn *et al.* (1890) showed that the prolonged action of 33 per cent. aqueous potassium hydroxide converts ecgonine into ψ-ecgonine, and Findlay (1953) has found that cocaine gives ψ-ecgonine methyl ester by the action of sodium methoxide in hot methanol.

§23a. Tropacocaine, $C_{15}H_{19}O_2N$, m.p. 49°, occurs in Java coca leaves. When heated with barium hydroxide solution, tropacocaine is hydrolysed to ψ-tropine and benzoic acid; thus the alkaloid is benzoyl-ψ-tropine.

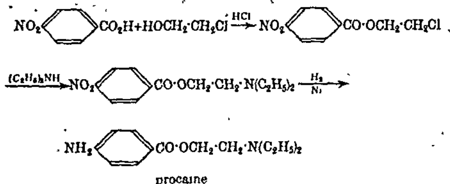


§23b. Cocaine substitutes. Cocaine is a very good local anæsthetic, but has certain disadvantages. The anæsthetic properties are lost if either the benzoyl group or the methyl ester group is removed; removal of the *N*-methyl group has no effect. A number of synthetic drugs have now been introduced to replace cocaine as a local anæsthetic; their anæsthetic properties are as good as those of cocaine, and they are less toxic. Two important substitutes are β-eucaine and procaine (novocaine).

β-Eucaine has been synthesised by treating acetone with ammonia and then treating the product, diacetonamine (see Vol. I), with diethyl acetal. The piperidone thereby produced is then reduced and finally benzoylated to give *β*-eucaine.



Procaine has been synthesised from *p*-nitrobenzoic acid.

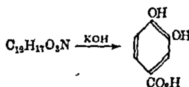


procaine

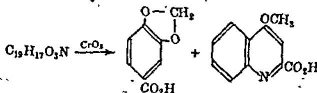
QUINOLINE GROUP

§24. Angostura alkaloids. A number of alkaloids have been isolated from angostura bark, *e.g.*, cusparine, galipine, galipoline, etc.

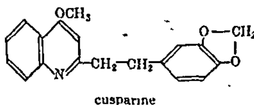
Cusparine, $C_{19}H_{17}O_3N$, m.p. 90–91°, has been shown to contain one methoxyl group (Zeisel method), and when fused with potassium hydroxide, protocatechuic acid is obtained.



On the other hand, controlled oxidation of cusparine gives piperonylic acid and 4-methoxyquinoline-2-carboxylic acid.

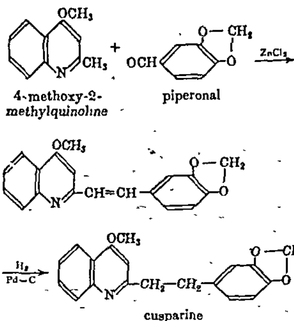


Consideration of this information led to the suggestion of the following structure for cusparine.

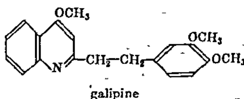


cusparine

This has been confirmed by synthesis (Späth *et al.*, 1924).

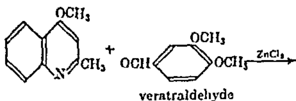


Galipine, $C_{20}H_{21}O_3N$, m.p. 113°, contains three methoxyl groups (Zeisel method). When oxidised with chromic acid, galipine produces 4-methoxyquinoline-2-carboxylic acid and veratric acid. Thus the formula of the alkaloid is probably:

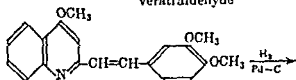


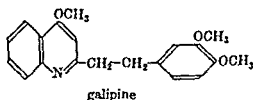
galipine

This has been confirmed by synthesis (Späth *et al.*, 1924).



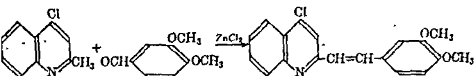
veratraldehyde

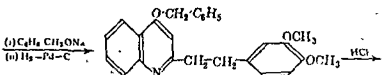


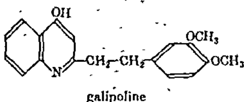


galipine

Galipoline, $C_{19}H_{19}O_3N$, m.p. 193°, contains two methoxyl groups and one phenolic group. When methylated with diazomethane, galipoline is converted into galipine. Thus one of the methoxyl groups in the latter is a hydroxyl group in the former. The position of this phenolic hydroxyl was shown to be in the quinoline nucleus by synthesis (Späth *et al.*, 1924).





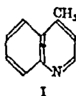


galipoline

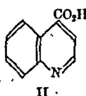
§25. Cinchona alkaloids. Cinchonine and quinine, together with many other alkaloids, occur in the bark of various species of *Cinchona*. Cinchonine may be regarded as the parent substance of the cinchona alkaloids, but quinine is the most important member of this group, its main use being in the treatment of malaria.

§25a. (+)-Cinchonine, $C_{19}H_{22}ON_2$, m.p. 264°, adds on two molecules of methyl iodide to form a di-quaternary compound; thus the alkaloid is a di-tertiary base. Since cinchonine forms a mono-acetate and a mono-benzoate, the molecule contains one hydroxyl group. Furthermore, this hydroxyl group is secondary alcoholic, since on oxidation, cinchonine forms the ketone *cinchoninone*. Cinchonine has been shown to contain one ethylenic double bond by the fact that it adds on one molecule of bromine or halogen acid, and that it is readily catalytically reduced, one molecule of hydrogen being added on.

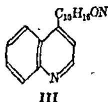
Fusion of cinchonine with potassium hydroxide gives lepidine (4-methylquinoline), I, and on vigorous oxidation with chromic acid in sulphuric acid solution, cinchoninic acid, II, is obtained (Königs, 1894). Thus cinchonine contains a quinoline nucleus with



I



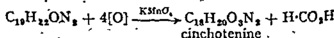
II



III

a side-chain in position 4 (III); this side-chain was referred to by Skraup as the "second-half" of the molecule. The hydroxyl group in cinchonine must be in this "second-half", since if it were not, then a hydroxy derivative or a carboxy derivative (since the hydroxyl is alcoholic) of cinchoninic acid would have been obtained.

Oxidation of cinchonine with permanganate gives cinchotenine and formic acid (Königs, 1879).

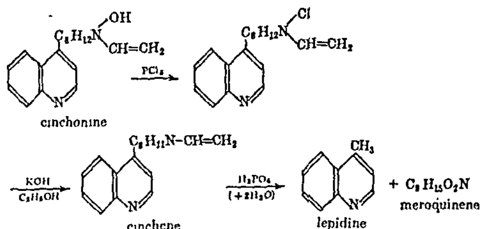


$$C_{19}H_{22}ON_2 + 4[O] \xrightarrow{KMnO_4} \underset{\text{cinchotenine}}{C_{18}H_{20}O_3N_2} + H{\cdot}CO_2H$$

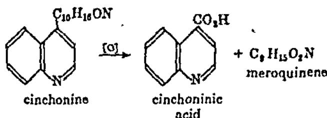
This suggests that there is a $-CH{=}CH_2$ group in the side-chain in the "second-half".

When treated with phosphorus pentachloride, followed by ethanolic potassium hydroxide, cinchonine is converted into cinchene which, when heated with 25 per cent. phosphoric acid, forms lepidine and a compound Königs named meroquinene (Königs *et al.*,

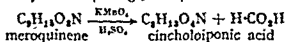
1884). With the information obtained so far, we may formulate the work of Königs as follows:



Meroquinene (meroquinenine) is also obtained, together with cinchoninic acid, when cinchonine is oxidised with chromic acid (Königs, 1894).



Thus the key to the structure of the "second-half" is the structure of meroquinene. The routine tests showed that meroquinene contains one carboxyl group and one double bond; the presence of the latter indicates that the $-CH{=}CH_2$ side-chain is still present in meroquinene. Oxidation of meroquinene with cold acid permanganate produces formic acid and cincholoiponic acid, the latter being a dicarboxylic acid (Königs, 1879). The formation of formic

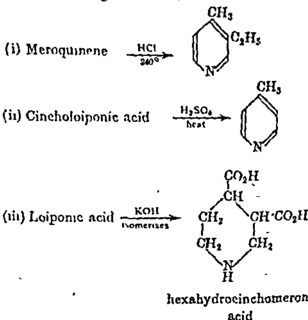


$$\underset{\text{meroquinene}}{C_9H_{15}O_2N} \xrightarrow[H_2SO_4]{KMnO_4} \underset{\text{cincholoiponic acid}}{C_8H_{13}O_4N} + H{\cdot}CO_2H$$

acid confirms the presence of the $-CH{=}CH_2$ side-chain in meroquinene. The presence of this group has also been demonstrated by the ozonolysis of meroquinene; formaldehyde is produced (Seekles, 1923). Oxidation of cincholoiponic acid with acid permanganate produces loiponic acid, $C_7H_{11}O_4N$ (Königs, 1890). This is also a dicarboxylic acid, and since it contains one methylene

group less than its precursor cincholoiponic acid, this suggests that the latter contains at least a side-chain $—CH_2·CO_2H$.

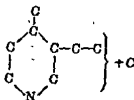
The reactions of the above three acids indicated that they were all secondary bases; that they all contained a piperidine ring is shown by the following reactions.



hexahydrocinchomeronic acid

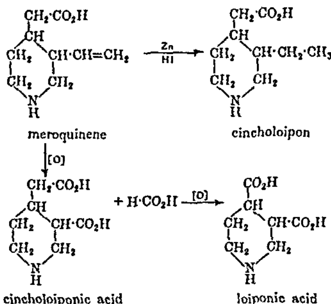
The structure of hexahydrocinchomeronic acid is known from its synthesis (*cf.* §21).

Consideration of the above results shows that a possible skeleton structure of meroquinene is:

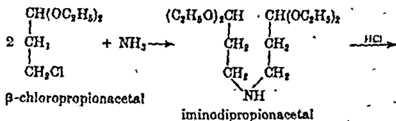


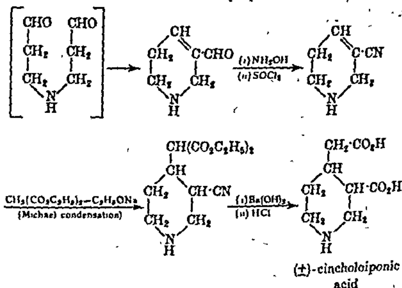
The problem then is to find the position of the remaining carbon atom. This carbon atom cannot be an *N*-methyl group, since all three acids are secondary bases. As we have seen, meroquinene contains a $—CH{=}CH_2$ group in the side-chain. A possible position for the extra carbon atom is the side-chain containing this unsaturated group, *i.e.*, the side-chain is an allyl group, $—CH_2·CH{=}CH_2$. All the foregoing facts can be explained on this basis, but the following fact cannot, *viz.*, that reduction of meroquinene gives

cincholoipon, $C_8H_{15}O_2N$, a compound which contains one carboxyl group and one *ethyl group*. Thus the unsaturated side-chain cannot be allyl (this should have given a propyl group on reduction); the side-chain is therefore vinyl. This leaves only one possible position for the extra carbon atom, *viz.*, 4; this would give a $-CH_2 \cdot CO_2H$ group at this position, and the presence of such a group has already been inferred (see above). All the reactions of meroquinene can therefore be explained on the following structures:



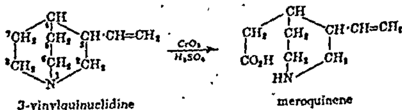
This formula for meroquinene is supported by the synthesis of cincholoiponic acid (Wohl *et al.*, 1907; *cf.* §17).



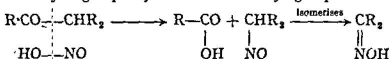


The racemic cincholoiponic acid was acetylated, and then this derivative was resolved by means of brucine; the (+)-form was identical with the acid obtained from meroquinene.

Since meroquinene is obtained from cinchonine by oxidation, the carbon atom of the carboxyl group in meroquinene will be the point of linkage to the "quinoline-half" at which scission of the "second-half" occurs. Since cinchonine is a di-tertiary base, the "second-half" therefore contains a tertiary nitrogen atom. But meroquinene is a *secondary* base, and it therefore follows that in its formation the tertiary nitrogen atom is converted into a secondary nitrogen atom, *a carboxyl group also being produced at the same time.* A possible explanation for this behaviour is that the tertiary nitrogen atom is a part of a bridged ring, one C—N bond being broken when cinchonine is oxidised:



Thus, in cinchonine, the "quinoline-half" must be joined *via* its side-chain at position 4 to the "quinuclidine-half" at position 8. The remaining problem is to ascertain the position of the secondary alcoholic group in the "second-half". Rabe *et al.* (1906, 1908) converted cinchonine into the ketone cinchoninone by gentle oxidation (chromium trioxide). This ketone, in which both nitrogen atoms are still tertiary, on treatment with amyl nitrite and hydrogen chloride, gives cinchoninic acid and an oxime. The formation of an acid and an oxime indicates the presence of the group —CO·CH—, *i.e.*, a methyne group adjacent to a carbonyl group:



The structure of the oxime obtained from cinchoninone was shown to be 8-oximino-3-vinylquinuclidine by its hydrolysis to hydroxylamine and meroquinene. If we assume that the secondary alcoholic group connects the "quinoline-half" to the quinuclidine nucleus, then the foregoing reactions may be written as follows, on the assumption that the structure of cinchonine is as given.

CHOH— $\xrightarrow{CrO_3}$ CO—

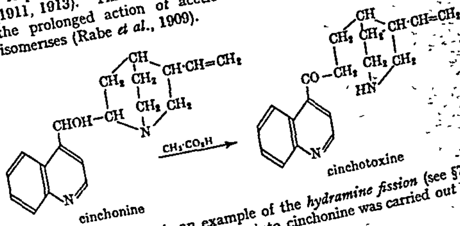
cinchonine cinchoninone

$\xrightarrow[HCl]{C_5H_{11}ONO}$ cinchoninic acid + oxime (HON=C) $\xrightarrow{acid}$

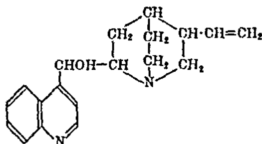
amide (O=C) $\xrightarrow{acid}$ CO_2H … HN

The above structure of cinchonine contains four dissimilar asymmetric carbon atoms, *viz.*, 3, 4, 8, and the carbon atom of the CHOH group (see 3-vinylquinuclidine for numbering). One pair of enantiomorphs is (±)-cinchonine, and another pair is (±)-*cinchonidine*; the configurations of C_3 and C_4 are the same in both, since both give the *same* 8-oximino-3-vinylquinuclidine.

A partial synthesis of cinchonine has been carried out by Rabe (1911, 1913). This starts from cinchotoxine, which is prepared by the prolonged action of acetic acid on cinchonine; the latter isomerises (Rabe *et al.*, 1909).

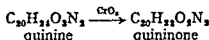


This isomerisation is an example of the *hydramine fission* (see §7). The conversion of cinchotoxine into cinchonine was carried out as follows:

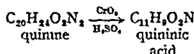


(±)-cinchonine

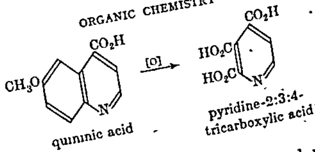
§25b. (−)-Quinine, $C_{20}H_{24}O_2N_2$, m.p. 177°, is used as a febrifuge and as an antimalarial. Since quinine adds on two molecules of methyl iodide to form a di-quaternary salt, it is therefore a di-tertiary base. *When heated with hydrochloric acid, quinine* eliminates one carbon atom as methyl chloride; therefore there is one methoxyl group present in the molecule. Since quinine forms a mono-acetate and a mono-benzoate, one hydroxyl group must be present, and that this is secondary alcoholic is shown by the fact that oxidation of quinine with chromium trioxide produces quininone, a ketone.



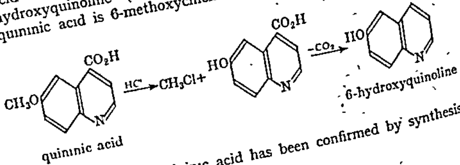
Quinine also contains one ethylenic double bond, as is shown by the fact that it adds on one molecule of bromine, etc. (*cf.* cinchonine). Oxidation of quinine with chromic acid produces, among other products, quininic acid.



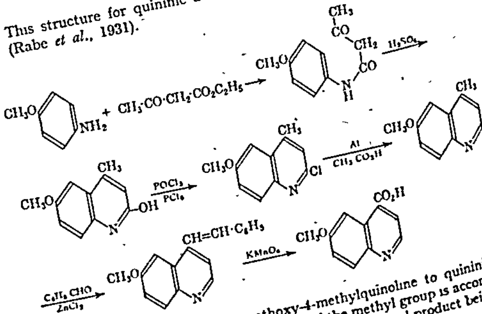
On the other hand, controlled oxidation of quinine with chromic acid gives quininic acid and meroquinene. Thus the "second-half" in both quinine and cinchonine is the same, and so the problem is to elucidate the structure of quininic acid. When heated with soda-lime, quininic acid is decarboxylated to a methoxyquinoline, and since, on oxidation with chromic acid, quininic acid forms pyridine-2.3:4-tricarboxylic acid, the methoxyl group must be a substituent in the benzene ring (of quinoline), and the carboxyl group at position 4 (Skraup, 1881). The position of the methoxyl group was ascertained by heating quininic acid with hydrochloric



acid and then decarboxylating the demethylated product; 6-hydroxyquinoline (a known compound) was obtained. Thus quininic acid is 6-methoxycinchoninic acid.

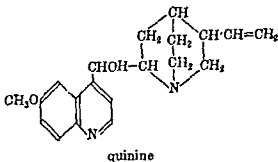


This structure for quininic acid has been confirmed by synthesis (Rabe *et al.*, 1931).



The direct oxidation of 6-methoxy-4-methylquinoline to quininic acid is extremely difficult; oxidation of the methyl group is accompanied by the oxidation of the benzene ring, the final product being pyridine-2.3.4-tricarboxylic acid (see §26).

Thus, on the basis of the foregoing evidence, the structure of quinine is;

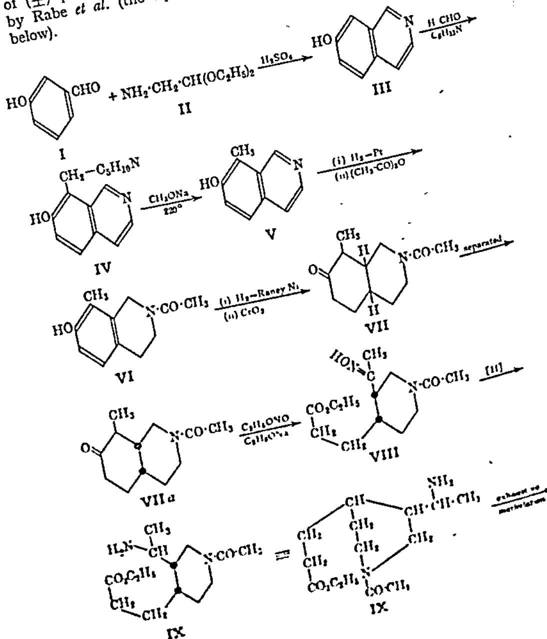


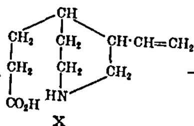
quinine

This formula contains the same four asymmetric carbon atoms as cinchonine; thus the same number of pairs of enantiomorphs is possible. One pair is (±)-quinine, and another pair is (±)-*quinidine*; the configurations of C_3 and C_4 are the same in quinine, quinidine, cinchonine and cinchonidine, since all four give the *same* 8-oximino-3-vinylquinuclidine (see §25a).

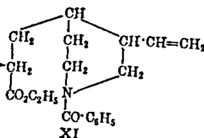
Rabe *et al.* (1918) carried out a partial synthesis of quinine starting from quinotoxine, which is prepared by heating quinine in acetic acid (*cf.* cinchotoxine). Woodward and Doering (1944) have synthesised (+)-quinotoxine, and so we now have a *total* synthesis of quinine. The following is Woodward and Doering's work up to (+)-quinotoxine, and from this to quinine is Rabe's work. *m*-Hydroxybenzaldehyde (I) is condensed with aminoacetal (II) and the product, 7-hydroxy*iso*quinoline (III), is treated with formaldehyde in methanol solution containing piperidine. The complex formed (IV) is converted into 7-hydroxy-8-methyl*iso*quinoline (V) by heating with methanolic sodium methoxide at 220°. V, on catalytic reduction (platinum) followed by acetylation, gives *N*-acetyl-7-hydroxy-8-methyl-1:2:3:4-tetrahydro*iso*quinoline (VI), which, on further catalytic reduction by heating with a Raney nickel catalyst under pressure and then followed by oxidation with chromium trioxide, is converted into *N*-acetyl-7-keto-8-methyldecahydro*iso*quinoline (VII). VII is a mixture of *cis*- and *trans*-isomers; these were separated and the *cis*-isomer (VII*a*; see §11 vii. IV for conventions) then treated with ethyl nitrite in the presence of sodium ethoxide to give the homomeroquinene derivative VIII. This, on reduction, gives IX, which may now be written more conveniently as shown. Exhaustive methylation of IX, followed by hydrolysis, gives *cis*-(±)-homomeroquinene (X). X, after esterification and benzoylation, gives XI which, on condensation with ethyl quininate (XII), produces XIII. This, on heating with 10 per

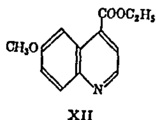
cent. hydrochloric acid, is hydrolysed and decarboxylated to (±)-quinotoxine (XIV). This was resolved *via* its dibenzoyltartrate (tartaric acid proved unsuccessful for resolution). The conversion of (±)-quinotoxine into quinine had already been accomplished by Rabe *et al.* (the equations for this conversion are also given below).

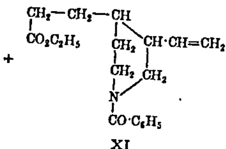



X

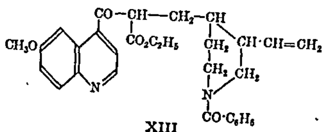

(i) C₂H₅OH–HCl
(ii) C₆H₅·COCl


XI

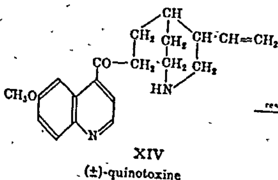

XII


XI


C₂H₅ONa


XIII


HCl


XIV
(±)-quinotoxine


resolved
(+)-isomer
(i) NaOBr
(ii) NaOH

(+)-quininone $\xrightarrow[C_2H_5ONa-C_2H_5OH]{Al}$

(±)-quinine $\xrightarrow{\text{resolved}}$ (−)-quinine

ISOQUINOLINE GROUP

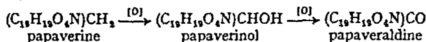
Opium alkaloids. Many alkaloids have been isolated from opium, and they are divided into two groups according to the nature of their structure:

(i) *isoQuinoline group, e.g.*, papaverine, laudanosine, etc.
(ii) *Phenanthrene group, e.g.*, morphine (see §27).

§26. Papaverine, $C_{20}H_{21}O_4N$, m.p. 147°, is one of the optically inactive alkaloids; it does not contain any asymmetric carbon atom. The structure of papaverine was established by Goldschmiedt and his co-workers (1883–1888). Since papaverine adds on one molecule of methyl iodide to form a quaternary iodide, the nitrogen atom in the molecule is in the tertiary state. The application of the Zeisel method shows the presence of four methoxyl groups: the demethylated product is known as papaveroline.

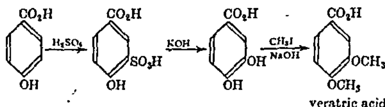
$$\underset{\text{papaverine}}{C_{20}H_{21}O_4N} + 4HI \longrightarrow 4CH_3I + \underset{\text{papaveroline}}{C_{16}H_{13}O_4N}$$

When oxidised with cold dilute permanganate, papaverine is converted into the secondary alcohol papaverinol, $C_{20}H_{21}O_5N$. This, on more vigorous oxidation with hot dilute permanganate, is oxidised to the ketone papaveraldine, $C_{20}H_{19}O_5N$ (it is the formation of this *ketone* that shows that papaverinol is a *secondary* alcohol). Finally, the prolonged action of hot permanganate oxidises papaveraldine to papaverinic acid, $C_{16}H_{13}O_7N$. This acid is a dibasic acid and still contains the keto group present in its precursor—it forms an oxime, etc.; papaverinic acid also contains two methoxyl groups. The foregoing reactions lead to the conclusion that papaverine contains a methylene group.



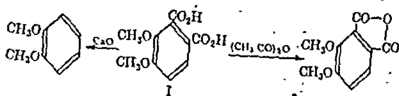
When oxidised with hot concentrated permanganate, papaverine (or the oxidised products mentioned above) is broken down into smaller fragments, *viz.*, veratric acid, metahemipinic acid, pyridine-2:3:4-tricarboxylic acid and 6:7-dimethoxy*iso*quinoline-1-carboxylic acid. Let us now consider the evidence for the structures of these compounds.

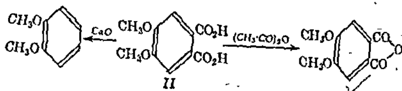
Veratric acid. When decarboxylated, veratric acid forms veratrole. Since this is *o*-dimethoxybenzene, veratric acid is therefore a dimethoxybenzoic acid. The position of the carboxyl group with respect to the two methoxyl groups (in the *ortho*-position) is established by the following synthesis.



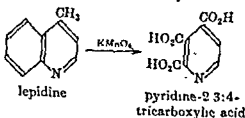
Thus veratric acid is 3·4-dimethoxybenzoic acid.

Metahemipinic acid. This is a dicarboxylic acid, and when decarboxylated by heating with calcium oxide, veratrole is formed; thus metahemipinic acid contains two methoxyl groups in the *ortho*-position. Furthermore, since the acid forms an anhydride when heated with acetic anhydride, the two carboxyl groups must be in the *ortho*-position. Thus metahemipinic acid is either I or II. Now metahemipinic acid forms only *one* monoester; II permits the formation of only one monoester, but I can give rise to two different monoesters. Thus II is metahemipinic acid; I is actually hemipinic acid (this isomer was known before metahemipinic acid).

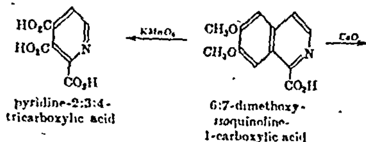


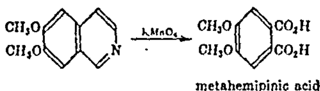


Pyridine-2:3:4-tricarboxylic acid. The routine tests showed that this contains three carboxyl groups, and since decarboxylation gives pyridine, the acid must be a pyridinetricarboxylic acid. The positions of the three carboxyl groups are established by the fact that this pyridinetricarboxylic acid is produced when lepidine (4-methylquinoline) is oxidised.



6:7-Dimethoxy*iso*quinoline-1-carboxylic acid. The usual tests showed that this compound contains one carboxyl group and two methoxyl groups. On oxidation, this acid forms pyridine-2:3:4-tricarboxylic acid; when decarboxylated, the acid forms a dimethoxy*iso*quinoline which, on oxidation, gives metahemipinic acid; thus the structure is established.



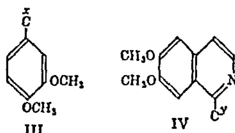


metahemipinic acid

We may now deduce the structure of papaverine as follows:

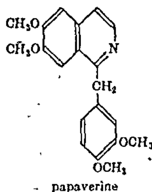
(i) The isolation of veratric acid indicates the presence of group III in papaverine.

(ii) The isolation of 6:7-dimethoxy*iso*quinoline-1-carboxylic acid indicates the presence of group IV in the molecule.



The presence of these two groups also accounts for the isolation of the other two fragments.

(iii) The total carbon content of III (9 carbon atoms) and IV (12 carbon atoms) is 21 carbon atoms. But papaverine contains only 20. There is, however, a $-CH_2-$ group present, and if we assume that C^x and C^y are one and the same carbon atom, *viz.*, the carbon atom of the CH_2 group, then the following structure of papaverine accounts for all the facts:



papaverine

Thus, with this formula, we can formulate the oxidation of papaverine as follows:

CH_3O, CH_3O, N, CH_2, OCH_3, OCH_3 —[O]→ CH_3O, CH_3O, N, CHOH, OCH_3, OCH_3

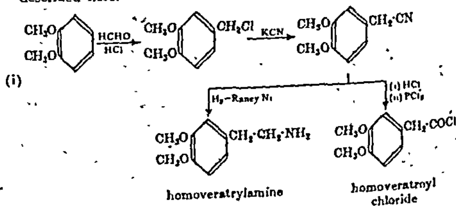
papaverine papaverinol

—[O]→ CH_3O, CH_3O, N, CO, OCH_3, OCH_3 —[O]→ HO_2C, HO_2C, N, CO, OCH_3, OCH_3

papaveraldine papaverinic acid

This structure for papaverine has been confirmed by synthesis. The first synthesis was by Pictet and Gams (1909), but Bide and Wilkinson (1945) carried out a simpler one, and it is this that is described here.

(i)



(ii)

heat

P_2O_5

$-H_2O$

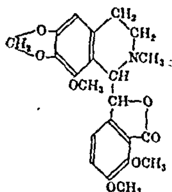
Pd on asbestos; 200°

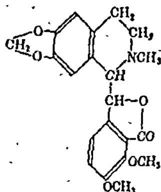
papaverine

§26a. Some other alkaloids of the *isoquinoline group* are :

laudanosine

laudanine


narcotine


hydrastine

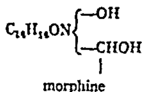
PHENANTHRENE GROUP

§27. Morphine, codeine and thebaine. These are three important opium alkaloids which contain the phenanthrene nucleus.

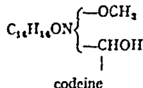
(—)-*Morphine*, $C_{17}H_{19}O_3N$, m.p. 254°, is the chief alkaloid in opium, and was the first alkaloid to be isolated (Sertürner, 1806). The usual tests show that the nitrogen atom is in the tertiary state, and since morphine forms a diacetate and a dibenzoate, two hydroxyl groups are therefore present in the molecule. Morphine gives the ferric chloride test for phenols, and dissolves in aqueous sodium hydroxide to form a *monosodium* salt, and this is reconverted into morphine by the action of carbon dioxide; thus *one* of the hydroxyl groups is phenolic (Matthiessen *et al.*, 1869). The second hydroxyl group is secondary alcoholic, as is shown by the following reactions. Halogen acids convert morphine into a monohalogeno derivative, one hydroxyl group being replaced by a halogen atom. When heated with methyl iodide in the presence of aqueous potassium hydroxide, morphine is methylated to give (—)-*codeine*, $C_{18}H_{21}O_3N$, m.p. 155° (Grimaux, 1881). Since codeine is no longer soluble in alkalis, it therefore follows that it is only the *phenolic* hydroxyl group in morphine that has been methylated. Furthermore, codeine can be oxidised by chromic acid to *codeinone*, a ketone (Hesse, 1884). Thus the hydroxyl group in codeine (and this one in morphine) is secondary alcoholic, and so codeine is the monomethyl (phenolic) ether of morphine.

(—)-*Thebaine*, $C_{19}H_{21}O_3N$, m.p. 193°, produces two molecules of

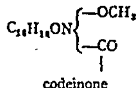
methyl iodide when heated with hydriodic acid (Zeisel method); hence thebaine is a dimethoxy derivative. When heated with sulphuric acid, thebaine eliminates one methyl group as methyl hydrogen sulphate, and forms codeinone (Knorr, 1906). The formation of a *ketone* led Knorr to suggest that thebaine is the methyl ether of the *enolic* form of codeinone. The foregoing work can thus be summarised by assigning the following formulae to the compounds described:



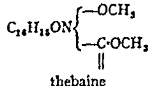
morphine



codeine



codeinone



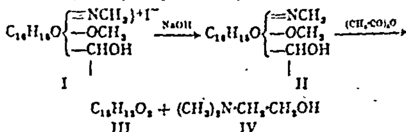
thebaine

So far, we have accounted for the functional nature of two of the oxygen atoms; the unreactivity of the third oxygen atom suggests that it is probably of the ether type (Vongerichten, 1881).

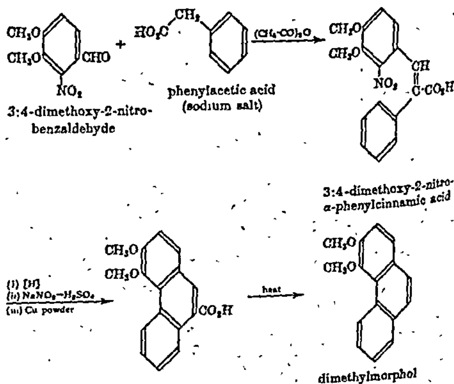
All three alkaloids are tertiary bases (each combines with one molecule of methyl iodide to form a methiodide). When heated with hydrochloric acid at 140° under pressure, morphine loses one molecule of water to form *apomorphine*, $C_{17}H_{17}O_2N$. Codeine, under the same conditions, also gives apomorphine (and some other products). Thebaine, however, when heated with dilute hydrochloric acid, forms *thebenine*, $C_{18}H_{19}O_3N$ (a secondary base), and with concentrated hydrochloric acid, morphothebaine, $C_{18}H_{19}O_3N$ (a tertiary base). Thus, in the formation of thebenine from thebaine, a tertiary nitrogen atom is converted into a secondary one. For this change to occur, the tertiary nitrogen must be of the type $>N{\cdot}R$, i.e., the nitrogen is in a ring system; had the nitrogen been in the group $—NR_2$, then the formation of a *primary base* would be expected.

When morphine is distilled with zinc dust, phenanthrene and a number of bases are produced (Vongerichten *et al.*, 1869). This suggests that a phenanthrene nucleus is probably present, and this has been confirmed as follows. When codeine methiodide is boiled with sodium hydroxide solution, α-methylmorphimethine is obtained and this, on heating with acetic anhydride

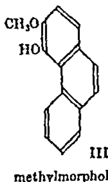
methylmorphol, III, and ethanoldimethylamine, IV (some of II isomerises to β-methylmorphimethine).



The structure of methylmorphol (III) was ascertained by heating it with hydrochloric acid at 180° under pressure; methyl chloride and a dihydroxyphenanthrene, *morphol*, were obtained. Oxidation of diacetylmorphol gives a diacetylphenanthraquinone; thus positions 9 and 10 are free. On further oxidation (permanganate), the quinone is converted into phthalic acid; therefore the two hydroxyl groups are in the same ring. Since the fusion of morphine with alkali gives protocatechuic acid, this shows that both hydroxyl groups in morphol are in the *ortho*-position. Finally, Pschorr *et al.* (1900) showed by synthesis that dimethylmorphol is 3:4-dimethoxyphenanthrene (*cf.* Pschorr synthesis, §2 *via.* X).



Then Pschorr *et al.* (1902) synthesised methylmorphol (III), and showed it to be 4-hydroxy-3-methoxyphenanthrene (in this synthesis Pschorr used 3-acetoxy-4-methoxy-2-nitrobenzaldehyde).

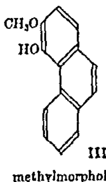


methylmorphol

The formation of ethanoldimethylamine (IV) from α-methylmorphimethine indicates that there is a $>NCH_3$ group in codeine (only *one* methyl iodide molecule adds to codeine to form codeine methiodide; it has also been shown above that this nitrogen is in a heterocyclic ring).

When β-methylmorphimethine is heated with water, the products obtained are trimethylamine, ethylene and *methylmorphenol* (Vongerichten, 1896). Demethylation of this compound with hydrochloric acid produces *morphenol*, a compound which contains one phenolic hydroxyl group and an inert oxygen atom. On fusion with potassium hydroxide, morphenol gives 3:4.5-trihydroxyphenanthrene (Vongerichten *et al.*, 1906). The structure of this compound was shown by the synthesis of 3:4:5-trimethoxyphenanthrene, which was found to be identical with the product obtained by methylating the trihydroxyphenanthrene obtained from morphenol (Pschorr *et al.*, 1912). Furthermore, the reduction of morphenol with sodium and ethanol gives morphol (Vongerichten, 1898). These results can be explained by assuming that morphenol has a structure containing an ether linkage in positions 4:5 (of the phenanthrene nucleus).

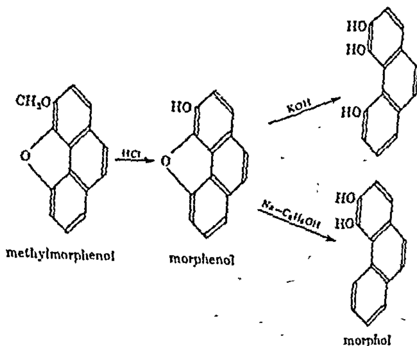
Then Pschorr *et al.* (1902) synthesised methylmorphol (III), and showed it to be 4-hydroxy-3-methoxyphenanthrene (in this synthesis Pschorr used 3-acetoxy-4-methoxy-2-nitrobenzaldehyde).



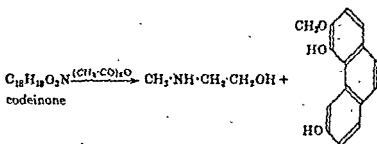
methylmorphol

The formation of ethanoldimethylamine (IV) from α-methylmorphimethine indicates that there is a $>NCH_3$ group in codeine (only *one* methyl iodide molecule adds to codeine to form codeine methiodide; it has also been shown above that this nitrogen is in a heterocyclic ring).

When β-methylmorphimethine is heated with water, the products obtained are trimethylamine, ethylene and *methylmorphenol* (Vongerichten, 1896). Demethylation of this compound with hydrochloric acid produces *morphenol*, a compound which contains one phenolic hydroxyl group and an inert oxygen atom. On fusion with potassium hydroxide, morphenol gives 3.4.5-trihydroxyphenanthrene (Vongerichten *et al.*, 1906). The structure of this compound was shown by the synthesis of 3:4:5-trimethoxyphenanthrene, which was found to be identical with the product obtained by methylating the trihydroxyphenanthrene obtained from morphenol (Pschorr *et al.*, 1912). Furthermore, the reduction of morphenol with sodium and ethanol gives morphol (Vongerichten, 1898). These results can be explained by assuming that morphenol has a structure containing an ether linkage in positions 4:5 (of the phenanthrene nucleus).

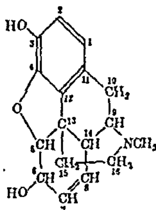


Codeinone, on heating with acetic anhydride, gives ethanolmethylamine and the diacetyl derivative of 4:6-dihydroxy-3-methoxyphenanthrene.

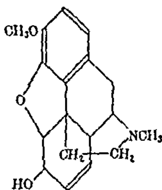


The position 3 of the methoxyl group and the position 4 of the hydroxyl group have already been accounted for; the hydroxyl group in the 6-position must therefore be produced from the oxygen of the keto group in codeinone.

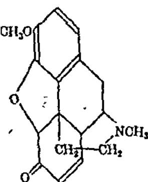
Based on the foregoing evidence, and a large amount of other experimental work, Gulland and Robinson (1923, 1925) have proposed the following structures.



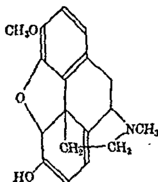
morphine



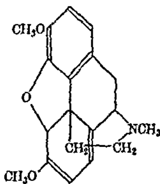
codeine



codeinone
(keto form)



codeinone
(enol form)



thebaine

Gates *et al.* (1956) have now synthesised morphine.

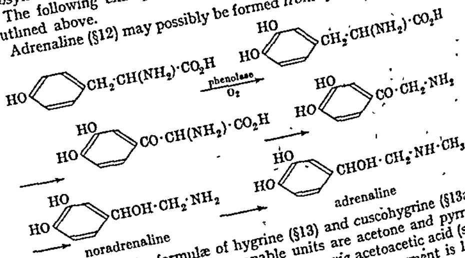
§28. **Biosynthesis of alkaloids.** As more and more structures of alkaloids were elucidated, it became increasingly probable that the precursors in the biosynthesis of alkaloids were amino-acids and amino-aldehydes and amines derived from them. A particularly interesting point is that the consideration of biosynthesis has led to deductions in structure, *e.g.*, Woodward (1948) proposed a biosynthesis of strychnine, and from this Robinson (1948) deduced the structure of emetine which was later confirmed by the synthetic work of Battersby *et al.* (1950).

We have already seen (§18. XIII) how keto-acids may be converted into amino-acids, and *vice-versa*. There are also enzymes which bring about the decarboxylation of amino-acids to amines and the decarboxylation of α-keto-acids to aldehydes. Thus amino-acids, amines and amino-aldehydes, together with formaldehyde

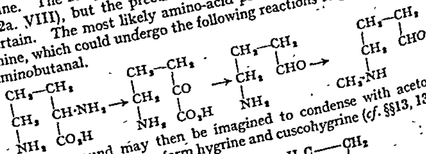
(or its equivalent) are believed to be the units involved in the biosynthesis of alkaloids.

The following examples of biosynthesis illustrate the principles outlined above.

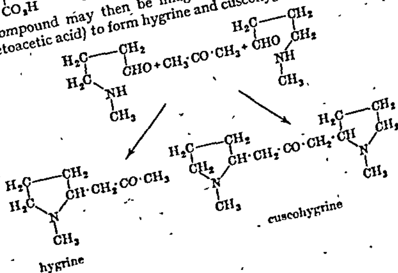
Adrenaline (§12) may possibly be formed from tyrosine as follows:



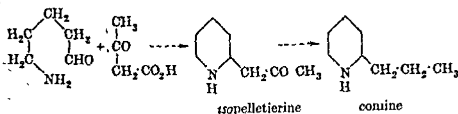
A study of the formulæ of hygrine (§13) and cuscohygrine (§13a) shows that the two most reasonable units are acetone and pyrrolidine. The biosynthesis of acetone occurs *via* acetoacetic acid (see §32a. VIII), but the precursor of the pyrrolidine fragment is less certain. The most likely amino-acid precursor appears to be ornithine, which could undergo the following reactions to give 4-methylaminobutanal.



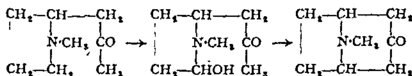
This compound may then be imagined to condense with acetone (or acetoacetic acid) to form hygrine and cuscohygrine (*cf.* §§13, 13a).



In the same way, the pelletierine group of alkaloids (§19) may all be imagined to be formed from 5-aminopentanal, *e.g.*; Anet *et al.* (1949) have condensed this aldehyde with acetoacetic acid at *p*H 11 to give *iso*pelletierine; and 5-methylaminopentanal with acetoacetic acid at *p*H 7 to give methyl*iso*pelletierine. The amino-acid precursor of 5-aminopentanal is most likely lysine (the homologue of ornithine). It should also be noted that conversion of the keto group in *iso*pelletierine into a methylene group gives coniine:

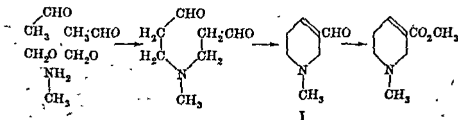


Now let us consider tropinone. Since this compound contains the hygrine skeleton, one possible mode of biosynthesis of tropinone could be *via* hygrine as the precursor:



On the other hand, tropinone has been synthesised from succinaldehyde, methylamine and acetonedicarboxylic acid under physiological conditions (§22). In this case, the problem is the nature of the precursor of succinaldehyde. Glutamic acid is one possibility, and succinic acid is another. The biosynthesis of cocaine (§23) is similar to that of tropinone.

The biosynthesis of some alkaloids containing a piperidine ring has already been discussed. Mannich (1942) has suggested that arecoline (§17) is formed as follows:

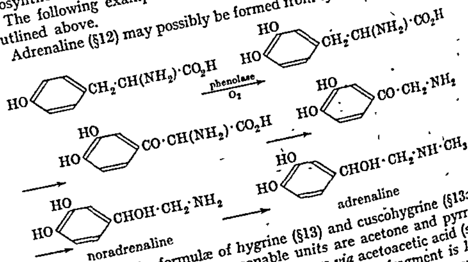


Mannich obtained I by carrying out the condensation with a mixture of acetaldehyde, formaldehyde and methylamine at room temperature at *p*H 3.

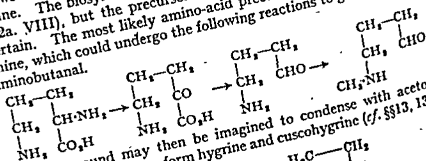
(or its equivalent) are believed to be the units involved in the biosynthesis of alkaloids.

The following examples of biosynthesis illustrate the principles outlined above.

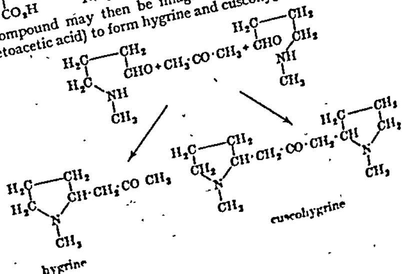
Adrenaline (§12) may possibly be formed from tyrosine as follows:



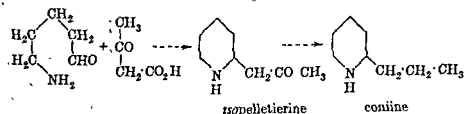
A study of the formulæ of hygrine (§13) and cuscohygrine (§13a) shows that the two most reasonable units are acetone and pyrrolidine. The biosynthesis of acetone occurs *via* acetoacetic acid (see §32a. VIII), but the precursor of the pyrrolidine fragment is less certain. The most likely amino-acid precursor appears to be ornithine, which could undergo the following reactions to give 4-methylaminobutanal.



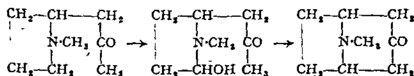
This compound may then be imagined to condense with acetone (or acetoacetic acid) to form hygrine and cuscohygrine (*cf.* §§13, 13a).



In the same way, the pelletierine group of alkaloids (§19) may all be imagined to be formed from 5-aminopentanal, *e.g.*; Anet *et al.* (1949) have condensed this aldehyde with acetoacetic acid at *p*H 11 to give *iso*pelletierine; and 5-methylaminopentanal with acetoacetic acid at *p*H 7 to give methyl*iso*pelletierine. The amino-acid precursor of 5-aminopentanal is most likely lysine (the homologue of ornithine). It should also be noted that conversion of the keto group in *iso*pelletierine into a methylene group gives coniine:

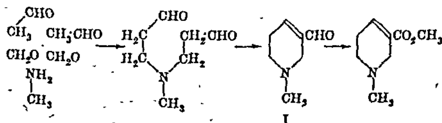


Now let us consider tropinone. Since this compound contains the hygrine skeleton, one possible mode of biosynthesis of tropinone could be *via* hygrine as the precursor:



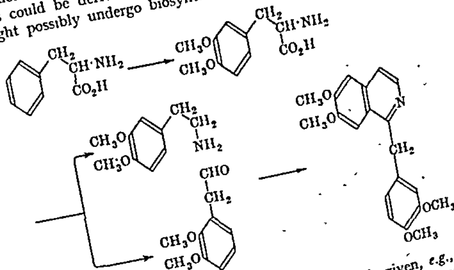
On the other hand, tropinone has been synthesised from succinaldehyde, methylamine and acetonedicarboxylic acid under physiological conditions (§22). In this case, the problem is the nature of the precursor of succinaldehyde. Glutamic acid is one possibility, and succinic acid is another. The biosynthesis of cocaine (§23) is similar to that of tropinone.

The biosynthesis of some alkaloids containing a piperidine ring has already been discussed. Mannich (1942) has suggested that arecoline (§17) is formed as follows:

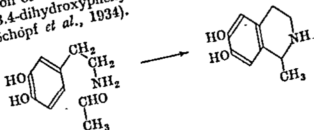


Mannich obtained I by carrying out the condensation with a mixture of acetaldehyde, formaldehyde and methylamine at room temperature at *p*H 3.

It has been suggested that the structural unit of *iso*quinoline alkaloids is phenylethylamine (and its oxygenated derivatives); this could be derived from phenylalanine, *e.g.*, papaverine (§26) might possibly undergo biosynthesis as follows:



Support for the plausibility of this mechanism is given, *e.g.*, by the formation of the tetrahydro*iso*quinoline from the condensation between 3.4-dihydroxyphenylethylamine and acetaldehyde at *p*H 3-5 (Schöpf *et al.*, 1934).



READING REFERENCES

Henry, *The Plant Alkaloids*, Churchill (1949, 4th ed.).

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). Vol. II. Ch. 15. Alkaloids.

Stewart, *Recent Advances in Organic Chemistry*, Longmans, Green (1931, 6th ed.). Vol. I. Chh. 11, 12. Alkaloids.

Stewart and Graham, *Recent Advances in Organic Chemistry* (1948, 7th ed.). Vol. II. Chh. 8, 9. Alkaloids.

Cook (Ed.), *Progress in Organic Chemistry*, Butterworth. Vol. I (1952). Ch. I. Molecular Structure of Strychnine, Brucine and Vomicine. Vol III (1955). Ch 5. Indole Alkaloids.

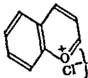
Manske and Holmes (Ed.), *The Alkaloids*, Academic Press. (Vol. I, 1950; —.)
Bergel and Morrison, Synthetic Analgesics, *Quart. Reviews (Chem. Soc.)*, 1948, **2**, 349.
Stern, Synthetic Approaches to the Morphine Structure, *Quart. Reviews (Chem. Soc.)*, 1951, **5**, 405
Gates and Tschudi, The Synthesis of Morphine, *J. Amer. Chem. Soc*, 1956, **78**, 1380.
McKenna, Steroidal Alkaloids, *Quart. Reviews (Chem. Soc.)*, 1953, **7**, 231.
Bentley, *The Chemistry of the Morphine Alkaloids*, Oxford Press (1954).
Bentley, *The Alkaloids*, Interscience Publishers (1957).
Wüest, A Hundred Years of Alkaloid Industry, *Chem. and Ind.*, **1937**, 1084.
Fodor *et al.*, The Configuration of Tropine and ψ-Tropine, *J.C.S.*, **1953**, 721.
Ann. Reports (Chem. Soc.), 1953, **50**, 164. Stereochemistry of the Tropines and Related Compounds.
Archer and Lewis, The Conformation of the Tropines and Some Related Compounds, *Chem. and Ind.*, **1954**, 853.
Glenn, The Structure of the Ergot Alkaloids, *Quart. Reviews (Chem. Soc.)*, 1954, **8**, 192.
Saxton, The Indole Alkaloids Excluding Harmine and Strychnine, *Quart. Reviews (Chem. Soc.)*, 1956, **10**, 108.
Sir Robert Robinson, *The Structural Relations of Natural Products*, Oxford Press (1955).
Wenkert, The Role of Oxidation in the Biogenesis of Alkaloids, *Experientia*, 1954, **10**, 346.
Morgan and Barltrop, Veratrum Alkaloids, *Quart. Reviews (Chem. Soc.)*, 1958, **12**, 34.

CHAPTER XV

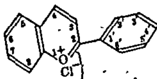
ANTHOCYANINS

§1. Introduction. Anthocyanins are natural plant pigments; they are glycosides and their aglycons, *i.e.*, the sugar-free pigments, are known as the *anthocyanidins*. The anthocyanins, which are water-soluble pigments, generally occur in the aqueous cell-sap, and are responsible for the large variety of colours in flowers; red—violet—blue. Willstätter *et al.* (1913–) showed that the various shades of colour exhibited by all flowers are due to a very small number of different compounds. Furthermore, these different compounds were shown to contain the same carbon skeleton, and differed only in the nature of the substituent groups. The anthocyanin pigments are amphoteric; their acid salts are usually red, their metallic salts usually blue and in neutral solution the anthocyanins are violet (see also §5).

§2. General nature of the anthocyanins. The fundamental nucleus in anthocyanidins is benzopyrylium chloride, but the parent compound is 2-phenylbenzopyrylium chloride or flavylium



benzopyrylium chloride



flavylium chloride

chloride. All anthocyanidins are derivatives of 3:5.7-trihydroxyflavylium chloride. The following table on page 659 shows some common anthocyanidins (as chlorides).

Various sugars have been found in anthocyanins; the most common are glucose, galactose and rhamnose, and the most important of these is glucose, which occurs as the diglucoside. Some pigments, as well as being glycosides, are also acylated derivatives, two common acids being *p*-hydroxybenzoic acid and malonic acid. The acid radical may be attached either to a phenolic hydroxyl group in the flavylium nucleus or to a hydroxyl group in the sugar residue.

Aglycon		Occurrence
Trivial name	Chemical name	
Pelargonidin	3:4′:5:7-Tetrahydroxyflavylium chloride	Present in orange-red to scarlet flowers, *e.g.*, scarlet *Pelargonium*, orange-red dahlia.
Cyanidin	3:3′:4′:5:7-Pentahydroxyflavylium chloride	Present in crimson to bluish-red flowers, *e.g.*, deep red dahlia, red roses, blue cornflower.
Delphinidin	3:3′:4:5:5′:7-Hexahydroxyflavylium chloride	Present in violet to blue flowers, *e.g.*, Delphinium.
Peonidin	3:4′:5:7-Tetrahydroxy-3′-methoxyflavylium chloride	Present in flowers less blue than the Cyanidin group, *e.g.*, red peony.
Malvidin (Syringidin)	3:4′:5.7-Tetrahydroxy-3′:5′-dimethoxyflavylium chloride	Present in flowers less blue than the Delphinidin group, *e.g.*, *Primula viscosa*.
Hirsutidin	3.4′:5-Trihydroxy-3′.5′:7-trimethoxyflavylium chloride	Present in *Primula hirsuta*.

A number of qualitative tests have been introduced to identify the various anthocyanins without actually isolating them (Robinson *et al.*, 1931–1933, 1938); *e.g.*,

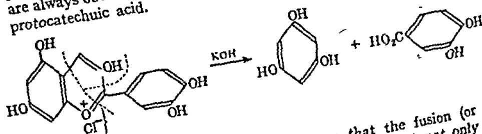
(i) The pigment is extracted with amyl (pentyl) alcohol in the presence of sodium acetate containing a trace of ferric chloride; cyanidin gives a blue colour, delphinidin a less intense blue colour, and the others still less colour or no colour at all.

(ii) A dilute sodium hydroxide solution of the pigment is shaken with air; delphinidin (and petunidin) is decolorised and the others are not.

(iii) More recently chromatographic analysis has been used to identify anthocyanins.

(iv) The spectra of the anthocyanins in the region 5000–5500 A are similar, but Geissman *et al.* (1953) have shown that the addition of aluminium chloride to solutions of certain anthocyanins shifts the absorption maximum. Only anthocyanins with the 3′:4′-dihydroxyl groups *free* show this shift, and so this observation may offer a method for analysing anthocyanin mixtures.

§3. Structure of the anthocyanidins. The anthocyanin is first hydrolysed with hydrochloric acid and the anthocyanidin is then isolated as the chloride. The usual analytical methods are applied to determine the number of hydroxyl and methoxyl groups present in the molecule. The structure of the anthocyanidin is ascertained by the nature of the products obtained by fusing the anthocyanidin with potassium hydroxide (Willstätter *et al.*, 1916); phloroglucinol or a methylated phloroglucinol and a phenolic acid are always obtained, *e.g.*, cyanidin chloride gives phloroglucinol and protocatechuic acid.

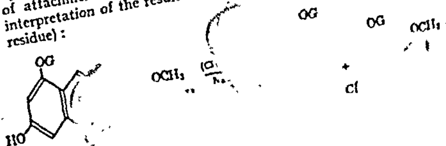


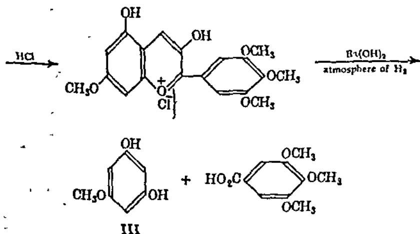
cyanidin chloride

This method suffers from the disadvantage that the fusion (or boiling with concentrated potassium hydroxide solution) not only degrades the anthocyanidin, but also often demethylates it at the same time. Thus the positions of the methoxyl groups in the original compound are now rendered uncertain. This difficulty was overcome by Karrer *et al.* (1927), who degraded the anthocyanidin with a ten per cent. solution of barium hydroxide or sodium hydroxide in an atmosphere of hydrogen; in this way, the methoxyl groups are left intact.

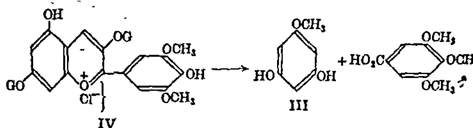
The next problem is to ascertain the positions of the sugar residues.

(i) Karrer *et al.* (1927) methylated the anthocyanin, then removed the sugar residues by hydrolysis (hydrochloric acid), and finally hydrolysed with barium hydroxide solution in an atmosphere of hydrogen; the positions of the *free* hydroxyl indicate the points of attachment of the sugar residues. In some cases, however, interpretation of the results is uncertain, *e.g.* (G represents a sugar residue):

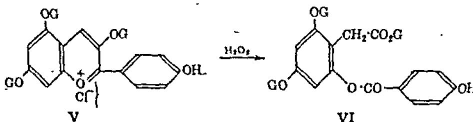




The problem is :- Which of the two hydroxyl groups in monomethylphloroglucinol was originally attached to G? The above results do not lead to a definite answer, since had the structure of the anthocyanin been IV instead of I, III would still have been obtained:



(ii) Hydrogen peroxide (15 per cent.) attacks anthocyanins as follows (Karrer *et al.*, 1927):



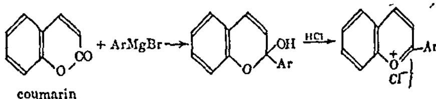
If the anthocyanin, V, has a glucose residue in the 3-position, then *this* glucose residue in VI is readily hydrolysed by dilute ammonia.

If the glucose residue in V is in either the 5- or 7-position, then this glucose residue in VI is removed only by heating with dilute hydrochloric acid. Thus position 3 can be distinguished from positions 5 or 7, but the latter two cannot be distinguished from each other.

(iii) Anthocyanins with a free hydroxyl group in the 3-position are very readily oxidised by ferric chloride; the anthocyanins are rapidly decolorised in this oxidation (Robinson *et al.*, 1931).

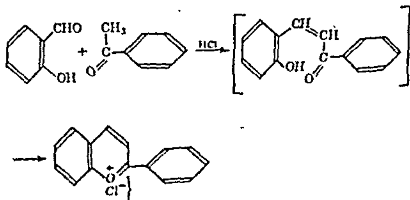
Conclusive evidence for the positions of the sugar residues is afforded by the synthesis of the anthocyanins (see, *e.g.*, cyanin, §5). In general, it has been found that glucose residues are linked at positions 3 or 3:5.

§4. General methods of synthesising the anthocyanidins.
(i) Willstätter (1914) synthesised anthocyanidins starting from coumarin.



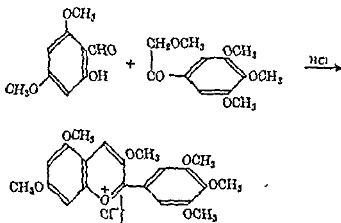
This method has very limited application.

(ii) Robinson has introduced a number of methods whereby *all* anthocyanidins can be prepared. The basic reaction of these methods is the condensation between *o*-hydroxybenzaldehyde and acetophenone in ethyl acetate solution which is then saturated with hydrogen chloride.



The original method of Robinson (1924) resulted in the formation

of a product in which the substituent groups were either all hydroxyl groups or all methoxyl groups, *e.g.*,



Robinson (1928, 1931) then modified this method so that the product could have both hydroxyl and methoxyl substituent groups, *e.g.*,

O·CO·C$_6$H$_5$ CHO HO OH + CH$_2$O·CO·CH$_3$ CO O·CO·CH$_3$ OCH$_3$ —HCl→

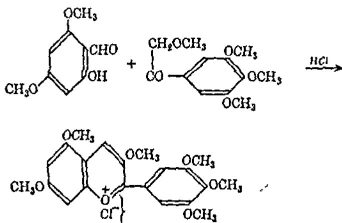
O·CO·C$_6$H$_5$ OH HO OH Cl$^-$ OCH$_3$ —NaOH (in N$_2$)→ OH OH HO OH CO OH OCH$_3$

—HCl (in N$_2$)→ OH OH HO OH Cl$^-$ OCH$_3$

peonidin chloride

The following is a brief account of the methods used by Robinson and his co-workers for preparing the substituted acetophenones and substituted benzaldehydes.

of a product in which the substituent groups were either all hydroxyl groups or all methoxyl groups, *e.g.*,



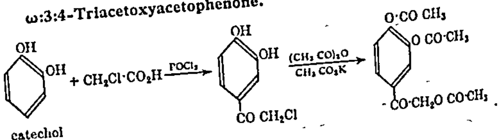
Robinson (1928, 1931) then modified this method so that the product could have both hydroxyl and methoxyl substituent groups, *e.g.*,

$$\xrightarrow{HCl} \qquad \xrightarrow[\text{(in } N_2)]{NaOH} \qquad \xrightarrow[\text{(in } N_2)]{HCl}$$

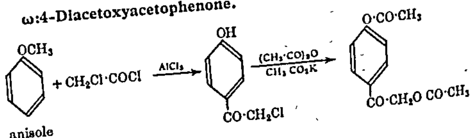
peonidin chloride

The following is a brief account of the methods used by Robinson and his co-workers for preparing the substituted acetophenones and substituted benzaldehydes.

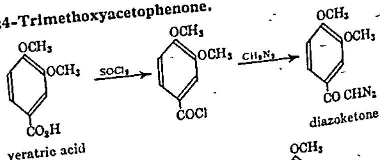
ω:3:4-Triacetoxyacetophenone.

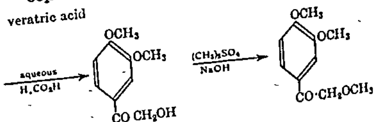


ω:4-Diacetoxyacetophenone.

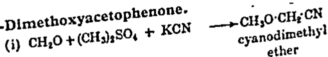


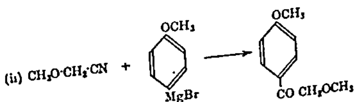
ω:3:4-Trimethoxyacetophenone.





ω:4-Dimethoxyacetophenone.





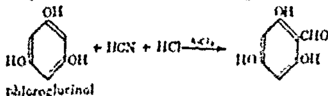
ω:3:4-Triacetoxy-5-methoxyacetophenone.

OH; HO, OH; CO_2H (gallic acid) $\xrightarrow[HCl]{CH_3OH}$ OH; HO, OH; CO_2CH_3 $\xrightarrow{(C_6H_5)_2CCl_2}$ $O—C(C_6H_5)_2—O$; HO; CO_2CH_3

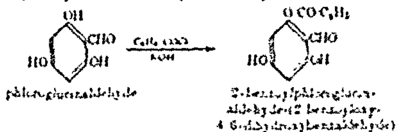
$\xrightarrow[NaOH]{(CH_3)_2SO_4}$ $O—C(C_6H_5)_2—O$; CH_3O; CO_2CH_3 $\xrightarrow{HCl}$ OH; CH_3O, OH; CO_2H $\xrightarrow[(2)\ SOCl_2]{(1)\ (CH_3\cdot CO)_2O}$

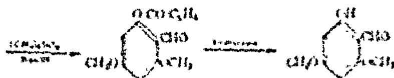
$O\ CO\cdot CH_3$; CH_3O, $O\ CO\ CH_3$; COCl $\xrightarrow[(2)\ CH_3\cdot CO_2H]{(1)\ CH_2N_2}$ $O\ CO\ CH_3$; CH_3O, $O\ CO\ CH_3$; $CO\cdot CH_2O\ CO\ CH_3$

2:4:6-Trihydroxybenzaldehyde (phloroglucinaldehyde).

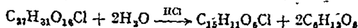


2-Hydroxy-4:6-dimethoxybenzaldehyde.



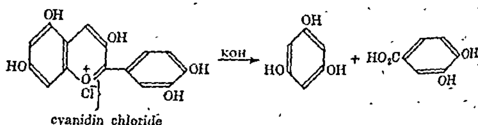


§5. Cyanidin chloride, $C_{15}H_{11}O_6Cl$. Cyanin chloride, on hydrolysis with hydrochloric acid, gives cyanidin chloride and two molecules of D-glucose.

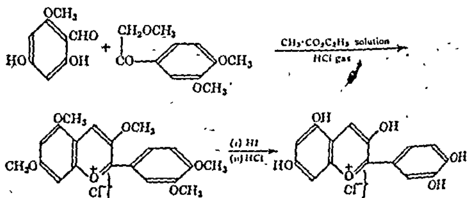


$$C_{27}H_{31}O_{16}Cl + 2H_2O \xrightarrow{HCl} C_{15}H_{11}O_6Cl + 2C_6H_{12}O_6$$

Since cyanidin chloride forms a penta-acetate, the molecule therefore contains five hydroxyl groups. No methoxyl groups are present, and so the potassium hydroxide fusion may be used to degrade this compound; this gives phloroglucinol and protocatechuic acid. Thus cyanidin chloride has the following structure:



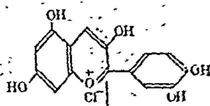
This structure has been confirmed by synthesis (Robinson *et al.*, 1928):



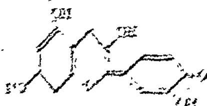
The formation of phloroglucinol and protocatechuic acid by the alkaline fusion of cyanidin chloride suggests a relationship to quercetin, since the latter also gives the same fusion products (see §14).

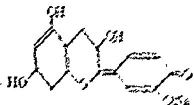
Cyanidin is insoluble in water, but is very soluble in ethanol. It is also soluble in aqueous sodium hydroxide, the solution being blue. The addition of hydrochloric acid changes the colour to purple when the solution is neutral, and when acid the solution

becomes red. According to Everest (1914), the colours are due to the following structures (see also Ch. XXXI, Vol. I):



Oxonium salt
Red in acid solution


Colour base
Purple in neutral solution



Salt of the colour base
Blue in alkaline solution

(ii) IV + $(CH_3\cdot COO)_4\cdot C_6H_7OBr$ (II) ⟶

IV: $CH_2OH\cdot CO\cdot C_6H_3(O\cdot CO\cdot CH_3)_2$

V: $CH_2O\cdot C_6H_7O(O\cdot CO\cdot CH_3)_4\cdot CO\cdot C_6H_3(O\cdot CO\cdot CH_3)_2$

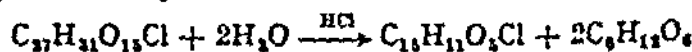
(iii) III + V $\xrightarrow{HCl}$ VI

VI: $O\cdot C_6H_{10}O_4\cdot O\cdot CO\cdot CH_3$; $O\cdot C_6H_7O(O\cdot CO\cdot CH_3)_4$; HO; Cl^-; $O\cdot CO\cdot CH_3$; $O\cdot CO\,CH_3$

$\xrightarrow{(i)\ KOH,\ (ii)\ HCl}$ VII

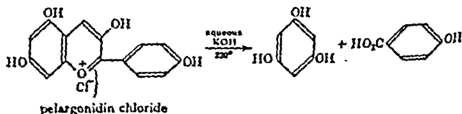
VII: $O\cdot C_6H_{11}O_5$; $O\cdot C_6H_{11}O_5$; HO; Cl^-; OH; OH

§6. **Pelargonidin chloride**, $C_{15}H_{11}O_5Cl$. This is formed, together with two molecules of glucose, when pelargonin chloride is hydrolysed with hydrochloric acid.



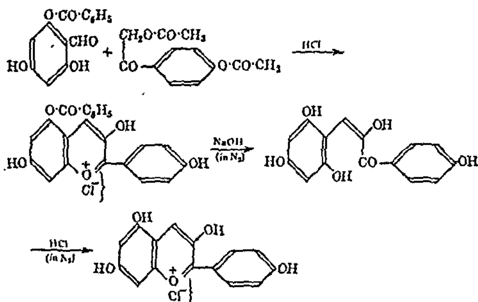
$$C_{27}H_{31}O_{15}Cl + 2H_2O \xrightarrow{HCl} C_{15}H_{11}O_5Cl + 2C_6H_{12}O_6$$

Since pelargonidin chloride forms a tetra-acetate, the molecule contains four hydroxyl groups. Furthermore, since there are no methoxyl groups present, the potassium hydroxide fusion or boiling with concentrated potassium hydroxide solution may be used to degrade the compound; the products are phloroglucinol and *p*-hydroxybenzoic acid, and so the structure is probably as shown:

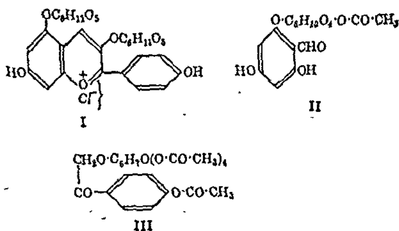


pelargonidin chloride

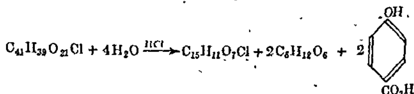
This structure has been confirmed by synthesis, *e.g.*, Robinson *et al.* (1928).



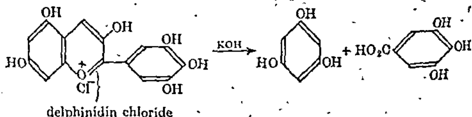
Pelargonin chloride, I, has been synthesised by Robinson *et al.* (1932) from 2-*O*-monoacetyl-β-glucosidylphloroglucinaldehyde, II, and ω-*O*-tetra-acetyl-β-glucosidoxy-4-acetoxyacetophenone, III (*cf.* cyanin chloride, §5).



§7. **Delphinidin chloride,** $C_{15}H_{11}O_7Cl$, is obtained, together with two molecules of glucose and two molecules of *p*-hydroxybenzoic acid, when delphinin chloride is hydrolysed with hydrochloric acid.

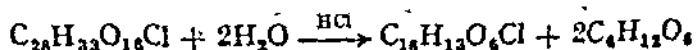


Delphinidin chloride contains six hydroxyl groups, and no methoxyl groups; on fusion with potassium hydroxide, the products are phloroglucinol and gallic acid.



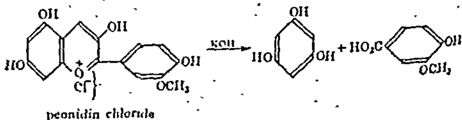
This structure has been confirmed by synthesis, starting from 2-benzoylphloroglucinaldehyde and ω-3:4:5-tetra-acetoxyacetophenone (Robinson *et al.*, 1930).

§8. **Peonidin chloride,** $C_{16}H_{13}O_6Cl$, is produced, together with two molecules of glucose, when peonin chloride is hydrolysed with hydrochloric acid.



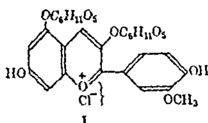
$$C_{28}H_{33}O_{16}Cl + 2H_2O \xrightarrow{HCl} C_{16}H_{13}O_6Cl + 2C_6H_{12}O_6$$

When heated with hydrogen iodide in the presence of phenol, peonidin chloride is demethylated to give cyanidin chloride. Thus *peonidin is the monomethyl ether of cyanidin.* Heating peonidin chloride with potassium hydroxide solution produces 4-hydroxy-3-methoxybenzoic acid and phloroglucinol. Thus:

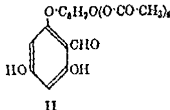


This structure has been confirmed by synthesis from 2-benzoylphloroglucinaldehyde and ω-4-diacetoxy-3-methoxyacetophenone (Robinson *et al.*, 1926).

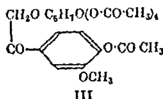
Peonin chloride, I, has been synthesised by Robinson *et al.* (1931), using 2-*O*-tetra-acetyl-β-glucosidylphloroglucinaldehyde, II, and ω-tetra-acetyl-β-glucosidoxy-4-acetoxy-3-methoxyacetophenone, III.



I

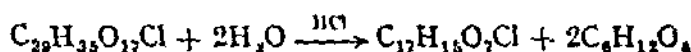


II



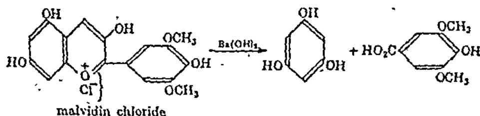
III

§9. **Malvidin chloride**, $C_{17}H_{15}O_7Cl$, is produced, together with two molecules of glucose, when malvin chloride is hydrolysed with hydrochloric acid.



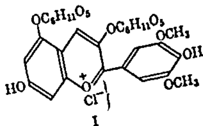
$$C_{29}H_{35}O_{17}Cl + 2H_2O \xrightarrow{HCl} C_{17}H_{15}O_7Cl + 2C_6H_{12}O_6$$

Malvidin chloride contains four hydroxyl groups and two methoxyl groups. When degraded by boiling barium hydroxide solution in an atmosphere of hydrogen, the products are phloroglucinol and *syringic acid* (4-hydroxy-3.5-dimethoxybenzoic acid). Thus:

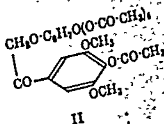


malvidin chloride

Robinson *et al.* (1928) confirmed this structure by synthesis, starting from 2-benzoylphloroglucinaldehyde and ω-acetoxy-4-benzyloxy-3:5-dimethoxyacetophenone (*cf.* §10). Robinson *et al.* (1932) have also synthesised **malvin chloride**, I, by condensing 2-*O*-tetra-acetyl-β-glucosidylphloroglucinaldehyde with ω-*O*-tetra-acetyl-β-glucosidoxy-4-acetoxy-3:5-dimethoxyacetophenone, II.

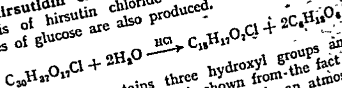


I



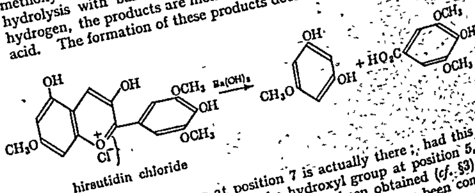
II

§10. Hirsutidin chloride, $C_{18}H_{17}O_7Cl$, is produced by the hydrolysis of hirsutin chloride with hydrochloric acid; two molecules of glucose are also produced.



$$C_{30}H_{27}O_{17}Cl + 2H_2O \xrightarrow{HCl} C_{18}H_{17}O_7Cl + 2C_6H_{12}O_6$$

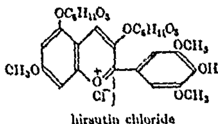
Hirsutidin chloride contains three hydroxyl groups and three methoxyl groups. Its structure is shown from the fact that on hydrolysis with barium hydroxide solution in an atmosphere of hydrogen, the products are monomethylphloroglucinol and syringic acid. The formation of these products does not prove conclusively



hirsutidin chloride

that the methoxyl group at position 7 is actually there; had this position been interchanged with the hydroxyl group at position 5, monomethylphloroglucinol would still have been obtained (*cf.* §3). The formula given for hirsutidin chloride, however, has been confirmed by synthesis, starting from 2-benzoyl-4-*O*-methylphloroglucinaldehyde and ω-acetoxy-4-benzyloxy-3:5-dimethoxyacetophenone (Robinson *et al.*, 1930).

Hirsutin chloride has also been synthesised by Robinson *et al.* (1932) from 2-*O*-tetra-acetyl-β-glucosidyl-4-*O*-methylphloroglucinaldehyde and ω-*O*-tetra-acetyl-β-glucosidoxy-4-acetoxy-3:5-dimethoxyacetophenone.

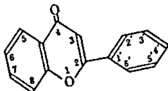


hirsutin chloride

FLAVONES

§11. Introduction. The flavones, which are also known as the **anthoxanthins**, are yellow pigments which occur in the plant kingdom. Flavones occur naturally in the free state, or as glycosides (the aglycon is the *anthoxanthidin* and the sugar is glucose or rhamnose), or associated with tannins. Chemically, the flavones are very closely related to the anthocyanins; the flavones are

hydroxylated derivatives of *flavone* (2-phenyl-4-chromone) which may be partially alkylated. In almost all cases positions 5 and 7



flavone

are hydroxylated, and frequently one or more of positions 3′, 4′, and 5′. The general method of ascertaining the structure of the flavones is similar to that used for the anthocyanins: the number of free phenolic groups and the number of methoxyl groups are first determined, and then the products obtained by alkaline fusion or hydrolysis are examined. Finally, the structure is confirmed by synthesis. Recently, Simpson *et al.* (1954) have shown that methoxyflavones may be demethylated selectively by hydrobromic acid, the relative rates being $3' > 4' > 7$. These authors have also shown that the relative rates of methylation of flavone-hydroxyl groups with methyl sulphate and sodium hydrogen carbonate in acetone solution are $7 > 4' > 3' > 3$. With methyl sulphate and aqueous alcoholic sodium carbonate, the exact reverse of this order is obtained. These results thus offer a method of ascertaining the positions of methoxyl groups in various methoxy-flavones.

§12. Flavone, $C_{15}H_{10}O_2$, occurs naturally as "dust" on flower leaves, etc. When boiled with concentrated potassium hydroxi solution, flavone, I, gives a mixture of four products, salicylic a (III), acetophenone (IV), *o*-hydroxyacetophenone (V) and benz acid (VI). The products, which are produced in the pairs III IV, and V and VI, arise from the fact that the opening of the py ring produces *o*-hydroxydibenzoylmethane, II, which then under scission in two different ways (II is a β-diketone).

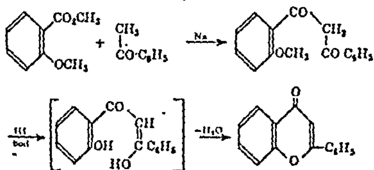
I $\xrightarrow{KOH}$ [$C_6H_4(OH)\cdot CO\cdot CH{=}C(OH)\cdot C_6H_5$] $\rightarrow$ $C_6H_4(OH)\cdot CO\cdot CH_2\cdot CO\,C_6H_5$ II

$\rightarrow$ $C_6H_4(OH)CO_2H$ (III) + $CH_3\cdot CO\cdot C_6H_5$ (IV)

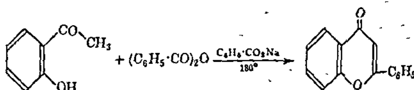
$\rightarrow$ $C_6H_4(OH)CO\cdot CH_3$ (V) + $C_6H_5\,CO_2H$ (VI)

In general, all the flavones give a mixture of four products when degraded with potassium hydroxide. The intermediate *o*-hydroxy-β-diketone can be isolated if *cold* alkali or an ethanolic solution of sodium ethoxide is used. On the other hand, if a normal solution of barium hydroxide is used as the degrading agent, then the products are usually salicylic acid and acetophenone (Simonis, 1917).

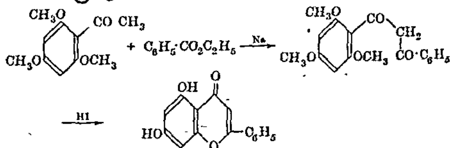
The structure given for flavone has been confirmed by synthesis. Many syntheses are known, *e.g.*, *the Kostanecki synthesis* (1900). This is a general method for synthesising flavones, and consists in condensing the ester of an alkylated salicylic acid with an acetophenone in the presence of sodium (this is an example of the Claisen condensation; this synthesis is a reversal of the formation of III and IV). Thus, for flavone itself, the reaction is carried out with methyl *o*-methoxybenzoate and acetophenone.



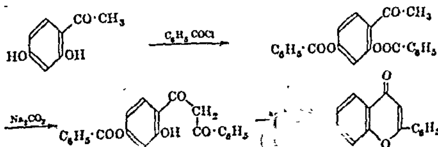
The most useful general synthetic method for preparing flavones is that of Robinson (1924). This is a reversal of the formation of V and VI; an *o*-hydroxyacetophenone is heated at about 180° with the anhydride and sodium salt of a substituted benzoic acid, *e.g.*, flavone:



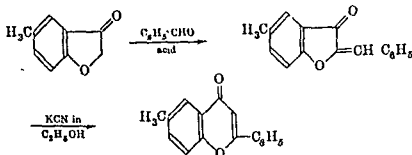
Another general method which is also a reversal of the formation of V and VI is illustrated by the preparation of *chrysin* (5:7-dihydroxyflavone) from 2:4:6-trimethoxyacetophenone and ethyl benzoate.



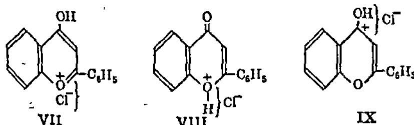
This preparation involves a Claisen condensation, and the following is also another general method which involves an "internal" Claisen condensation.



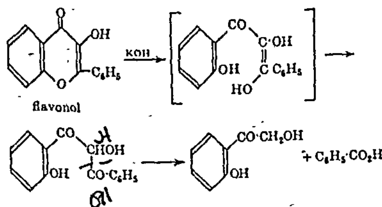
A recent metho ... ng flavo ... expansion of 2-benzylidenec ... (Wheel ... *e.g.*,



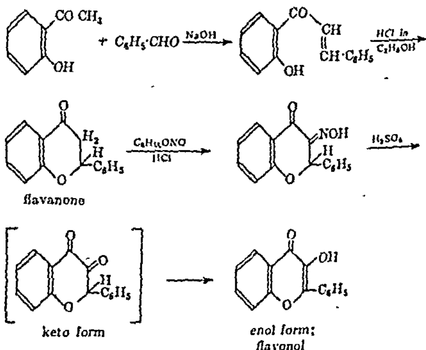
Most flavones are yellow solids which are soluble in water, ethanol and dilute acids and alkalis. The oxonium salts are usually more highly coloured than the free bases; the flavones do not occur ... flavone salts ... ıd according ... , IX appears to be most likely (Brown, 1951).



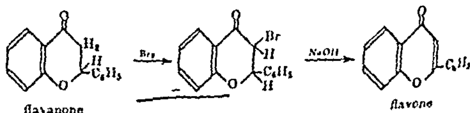
§13. ... Flavonol is widely distribı ... ɜ form of glycosides. Wheı ... ɔtassium hydroxide, flavonol gives *o*-hydroxybenzoylmethanol and benzoic acid. This suggests that flavonol is 3-hydroxyflavone (3-hydroxy-2-phenyl-γ-chromone).



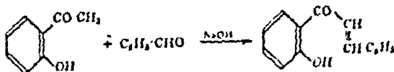
This structure has been confirmed by various syntheses, *e.g.*, Kostanecki *et al.* (1904). This is a general method, and uses the Claisen reaction between *o*-hydroxyacetophenones and substituted benzaldehydes, *e.g.*, flavonol.

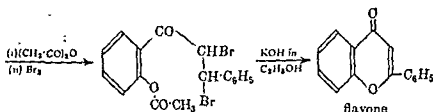


Thus synthesis, starting from flavanone, has been adapted to the preparation of flavones.

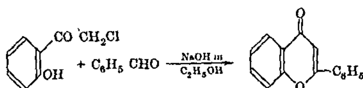


An alternative general method for preparing flavones based on the flavonol synthesis is as follows (Kostanecki *et al.*, 1898):

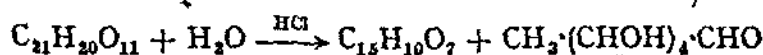




This synthesis has been simplified by Wheeler *et al.* (1955); these authors prepared flavones by condensing *ω*-chloro-*o*-hydroxyacetophenones with aromatic aldehydes in the presence of ethanolic sodium hydroxide, *e.g.*,

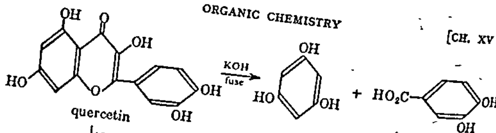


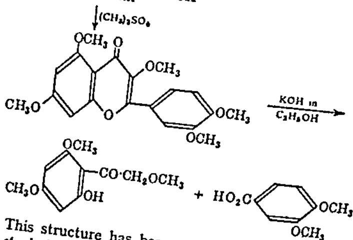
§14. **Quercetin**, $C_{15}H_{10}O_7$, occurs as the glycoside *quercitrin* in the bark of *Quercus tinctoria*; quercitrin appears to be the most widely distributed natural pigment. On hydrolysis with acids, quercitrin forms quercetin and one molecule of rhamnose.



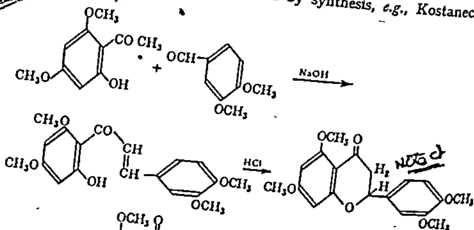
$$C_{21}H_{20}O_{11} + H_2O \xrightarrow{HCl} C_{15}H_{10}O_7 + CH_3\cdot(CHOH)_4\cdot CHO$$

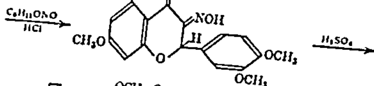
Quercetin contains five hydroxyl groups; no methoxyl groups are present; on fusion with potassium hydroxide, phloroglucinol and protocatechuic acid are obtained (*cf.* cyanidin, §5). Also, when quercetin is methylated and the product, pentamethylquercetin, boiled with an ethanolic solution of potassium hydroxide, 6-hydroxy-*ω*:2:4-trimethoxyacetophenone and veratric acid are obtained. These results suggest that quercetin is 3:3′.4′.5:7-pentahydroxyflavone.

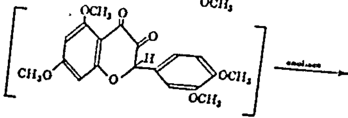


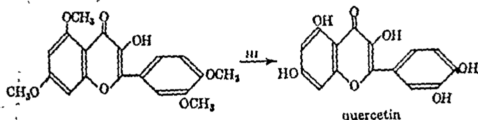


This structure has been confirmed by synthesis, *e.g.*, Kostanecki *et al.* (1904).





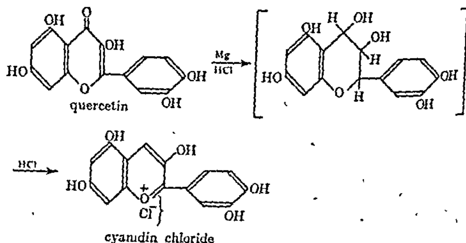




Another synthesis is that of Robinson *et al.* (1926); it is a general method for flavonols (*cf.* flavone, §12): ω-methoxyphloroacetophenone is condensed with veratric anhydride in the presence of the potassium salt of veratric acid.

The position of the rhamnose residue in quercitrin has been shown to be 3 (Herzig *et al.*, 1912).

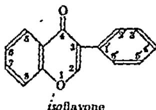
Before leaving this problem of quercetin, let us consider its relationship to cyanidin (§5). As we have seen, the relationship between the two compounds is suggested by the fact that both give the same products when fused with potassium hydroxide. Willstatter *et al.* (1914) reduced quercetin with magnesium in hydrochloric acid containing mercury, and thereby obtained a small amount of cyanidin chloride.



Bauer *et al.* (1954) have converted the penta-acetate of quercetin into cyanidin chloride by means of lithium aluminium hydride.

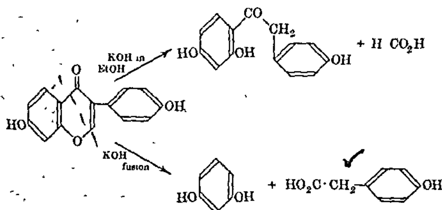
ISOFLAVONES

§14a. *iso*Flavones are hydroxylated derivatives of *iso*flavone (3-phenyl-4-chromone) which may be partially alkylated. The



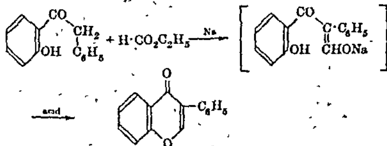
*iso*flavones occur naturally, but are not so widespread as the

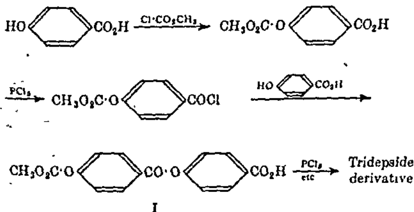
flavones; they occur either in the free state or as glycosides. The general method of ascertaining the structure of *iso*flavones is similar to that used for the flavones (see §§3, 11). Thus fusion with potassium hydroxide breaks down the molecule into two fragments, and hydrolysis with ethanolic potassium hydroxide permits the isolation of intermediates. This may be illustrated with *daidzein* (Walz, 1931):



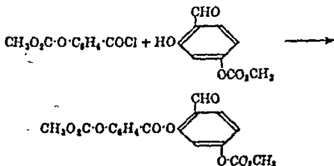
Oxidation with alkaline hydrogen peroxide may also be used in degrading *iso*flavones; recognisable fragments are not usually obtained by this method, but sometimes information may be obtained about the substituents in the 3-phenyl nucleus, *e.g.*, *genistein* (4′.5·7-trihydroxy*iso*flavone) gives *p*-hydroxybenzoic acid.

The final proof of the structure of an *iso*flavone lies in its synthesis. A general method of synthesising *iso*flavones is that of Späth *et al.* (1930); *e.g.*, *iso*flavone itself may be synthesised from benzyl *o*-hydroxyphenyl ketone and ethyl formate:





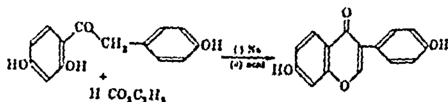
I may be hydrolysed to the didepside by means of cold alkali. By using different phenolic acids, it is possible to synthesise a large variety of depsides. When the hydroxyl group is *meta* or *para* to the carboxyl group, the phenolic acid is readily carboxymethylated, but *ortho*-hydroxyl groups are very resistant under the same conditions (spatial effect; see Vol. I). Reaction can, however, be brought about by condensing *o*-hydroxyacids with methyl chloroformate in the presence of a base, *e.g.*, dimethylaniline. There is also the further difficulty that *ortho*-hydroxyl groups do not react with acid chlorides (spatial effect). This has been overcome by condensing an acid chloride with an *o*-phenolic aldehyde, *e.g.*,



§16. Tannins. These are widely distributed in plants; many are glycosides. One of the best sources of tannin is nutgall. The tannins are colourless non-crystalline substances which form colloidal solutions in water; these solutions have an astringent taste. Tannins precipitate proteins from solution, and they form a bluish-black colour with ferric salts, a property which is used in the manufacture of ink. Tannins also precipitate many alkaloids from their solutions.

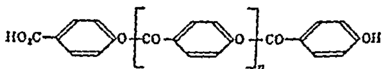
All tannins contain polyhydroxyphenols or their derivatives.

By using substituted ketones, various *isoflavones* may be synthesised, *e.g.*, daidzein from 2,4-dihydroxyphenyl *p*-hydroxybenzyl ketone (Wessely *et al.*, 1933):



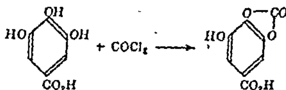
DEPSIDES

§15. Depsides. Phenolic acids, by the interaction of the carboxyl group of one molecule with the hydroxyl group of another, give rise to **depsides**:

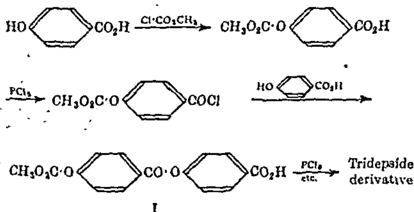


If n is zero, then the molecule is a didepside; if n is 1, then a tridepside; etc. The main sources of the depsides are the lichens.

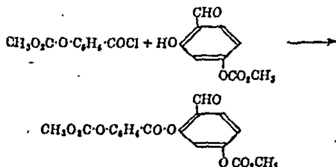
In order to synthesise depsides in a known fashion, it is necessary to protect hydroxyl groups. Fischer (1919) carried this out by means of acetylation (acetic anhydride) or by introducing a carbomethoxyl group (with methyl chloroformate); two hydroxyl groups in the *ortho*-position may be protected by means of carbonyl chloride, *e.g.*, gallic acid forms the following compound.



Let us consider the synthesis of a depside from a monohydroxybenzoic acid.



I may be hydrolysed to the didepside by means of cold alkali. By using different phenolic acids, it is possible to synthesise a large variety of depsides. When the hydroxyl group is *meta* or *para* to the carboxyl group, the phenolic acid is readily carboxymethylated, but *ortho*-hydroxyl groups are very resistant under the same conditions (spatial effect; see Vol. I). Reaction can, however, be brought about by condensing *o*-hydroxyacids with methyl chloroformate in the presence of a base, *e.g.*, dimethylaniline. There is also the further difficulty that *ortho*-hydroxyl groups do not react with acid chlorides (spatial effect). This has been overcome by condensing an acid chloride with an *o*-phenolic aldehyde, *e.g.*,



§16. Tannins. These are widely distributed in plants; many are glycosides. One of the best sources of tannin is nutgall. The tannins are colourless non-crystalline substances which form colloidal solutions in water; these solutions have an astringent taste. Tannins precipitate proteins from solution, and they form a bluish-black colour with ferric salts, a property which is used in the manufacture of ink. Tannins also precipitate many alkaloids from their solutions.

All *tannins contain polyhydroxyphenols or their derivatives.*

Some tannins are hydrolysable by acids, and others are not; those which can be hydrolysed by acid give variable yields of gallic acid.

Reading References

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). Vol II Ch 18. The Anthocyanins and the Flavones.

Stewart and Graham, *Recent Advances in Organic Chemistry*, Longmans, Green. Vol. II (1948, 7th ed.). (i) Ch. 10. Anthocyanins. (ii) Ch. 11. Depsides and Tannins.

Perkin and Everest, *The Natural Organic Colouring Matters*, Longmans, Green (1918).

Elderfield (Ed.), *Heterocyclic Compounds*, Wiley. Vol. II (1951). (i) Ch. 8. Chromones, Flavones and Isoflavones. (ii) Ch. 9. Chromenols, Chromenes, and Benzopyrylium Salts.

Hill, The Synthesis and Structure of Benzopyrylium (Chromylium) Salts, *Chem. Reviews*, 1936, **19**, 27.

Robinson, Natural Colouring Matters and their Analogues, *Chem. and Ind.*, **1933**, 737.

Robinson, Über die Synthese von Anthocyaninen, *Ber.*, 1934, **67A**, 85.

Robinson, Chemistry of the Anthocyanins, *Nature*, 1935, **135**, 732.

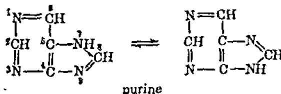
Warburton, The *iso*Flavones, *Quart. Reviews (Chem. Soc.)*, 1954, **8**, 67.

Jain and Seshadri, Nuclear Methylation of Flavones and Related Compounds, *Quart. Reviews (Chem. Soc.)*, 1956, **10**, 169.

CHAPTER XVI

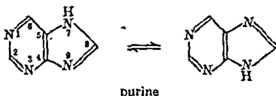
PURINES AND NUCLEIC ACIDS

§1. Introduction. Purine is the parent substance of a group of cyclic diureides; it does not occur naturally, but it was used by E. Fischer to name systematically the naturally occurring derivatives. Purine exists in two tautomeric forms, and its structure consists of a pyrimidine ring fused to a glyoxaline ring. In the earlier literature, the formula of purine was written as follows (the method of numbering is also shown):



purine

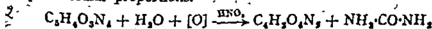
The tendency nowadays is to write the formula as follows:



purine

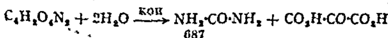
§2. Uric acid. Guano (birds' excrement found on islands near the western coast of South America) contains up to about 25 per cent. uric acid; about 90 per cent. of snakes' excrement is ammonium urate. Small amounts of uric acid are also present in human urine; it was first discovered by Scheele (1776) in urinary calculi.

Liebig and Wöhler (1834) showed that the molecular formula of uric acid is $C_5H_4O_3N_4$. These authors also found, in 1838, that the oxidation of uric acid with nitric acid gives alloxan and urea in equimolecular proportions.



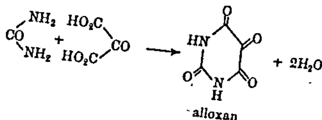
$$C_5H_4O_3N_4 + H_2O + [O] \xrightarrow{HNO_3} C_4H_2O_4N_2 + NH_2{\cdot}CO{\cdot}NH_2$$

Structure of alloxan, $C_4H_2O_4N_2$. When hydrolysed with alkali, alloxan produces one molecule of urea and one molecule of mesoxalic acid.

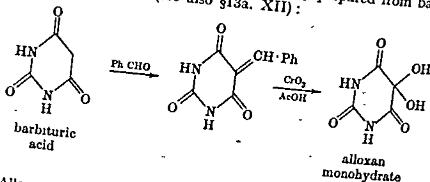


$$C_4H_2O_4N_2 + 2H_2O \xrightarrow{KOH} NH_2{\cdot}CO{\cdot}NH_2 + CO_2H{\cdot}CO{\cdot}CO_2H$$

Since alloxan contains no free amino or carboxyl groups, the products of hydrolysis suggest that alloxan is mesoxalylurea; this cyclic structure has been confirmed by the direct union of urea and mesoxalic acid to give alloxan (Liebig and Wöhler, 1838).



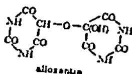
Alloxan, as its monohydrate, is conveniently prepared from barbituric acid as follows (see also §13a. XII):



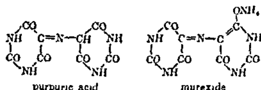
Alloxan is a strongly acid compound (in the *enol* form); it crystallises with four molecules of water of crystallisation. Three of these are readily lost on warming, but the fourth is lost only when the monohydrate is heated to 150°. Because of this, it is believed that the fourth molecule of water is not water of crystallisation but water of constitution (*cf.* chloral hydrate, Vol. I).

Alloxan stains the ski[illegible]

[illegible] only. Alloxantin is pro[illegible] alloxan (one molecule) with half a molecule of hydrogen sulphide, or by mixing aqueous solutions of alloxan and dialuric acid.



When heated with ammonia in ethanolic solution, alloxantin forms *murexide*, which is the ammonium salt of *purpuric acid* (an unstable compound).



Murexide is soluble in water, giving a purple solution which turns blue on the addition of alkali. Purpuric acid slowly hydrolyses in solution to form alloxan and uramil.

When uric acid is oxidised with an aqueous suspension of lead dioxide, the products are allantoin and carbon dioxide (Liebig and Wöhler, 1838). These products are obtained in quantitative yield if the oxidation is carried out with alkaline permanganate (Behrend, 1904).



$$C_5H_4O_3N_4 + H_2O + [O] \rightarrow C_4H_6O_3N_4 + CO_2$$

Structure of allantoin, $C_4H_6O_3N_4$ (Baeyer, 1861–1864). When hydrolysed with alkali, allantoin forms two molecules of urea and one molecule of glyoxylic acid.



$$C_4H_6O_3N_4 + 2H_2O \rightarrow 2NH_2{\cdot}CO{\cdot}NH_2 + CHO{\cdot}CO_2H$$

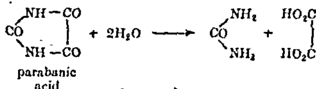
The formation of these hydrolytic products suggests that allantoin is the diureide of glyoxylic acid.

On oxidation with nitric acid, allantoin forms urea and *parabanic acid* in equimolecular proportions.

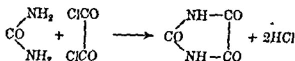


$$C_4H_6O_3N_4 + [O] \xrightarrow{HNO_3} NH_2{\cdot}CO{\cdot}NH_2 + C_3H_2O_3N_2$$

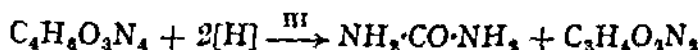
Now parabanic acid, on hydrolysis, gives urea and oxalic acid, and since there are no free amino or carboxyl groups present in the molecule, this suggests that parabanic acid is oxalylurea.



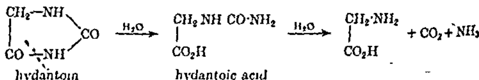
This structure has been confirmed by synthesis, *e.g.*, oxalyl chloride condenses with urea to form parabanic acid (Bornwater, 1912).



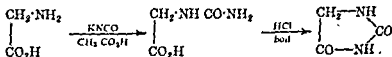
Thus, from the above facts, it can be seen that allantoin contains the parabanic acid nucleus joined to a molecule of urea. The point of the attachment is deduced from the following experimental evidence. When reduced with concentrated hydriodic acid at 100°, allantoin forms urea and *hydantoin*.



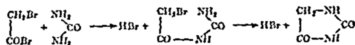
Hydantoin, on controlled hydrolysis, gives *hydantoic acid* (*ureido-acetic acid*) and this, on further hydrolysis, gives glycine, ammonia and carbon dioxide. These results suggest that hydantoin is glycollylurea



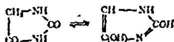
This structure for hydantoin has been confirmed by synthesis, *e.g.*, West (1918).



Hydantoin, m.p. 216°, may also be prepared by the electrolytic reduction of parabanic acid, or by the action of bromoacetyl bromide on urea.

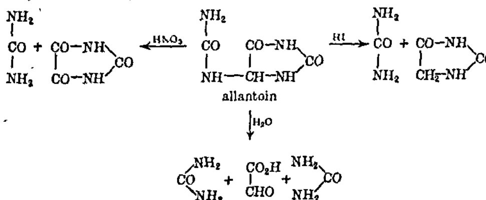


Hydantoin behaves as a tautomeric substance; the enol form is acidic and forms salts.

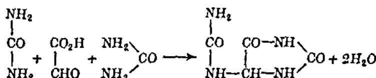


Hydantoin is oxidised to parabanic acid by bromine water.

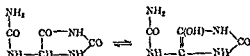
Thus the following structure for allantoin would account for all of the foregoing results:



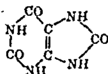
This has been confirmed by synthesis by heating urea with glyoxylic acid at 100° (Grimaux, 1876).



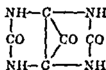
Examination of the structure of allantoin shows that it contains an asymmetric carbon atom; hence two optically active forms are possible. Both forms have been obtained, and they have been found to racemise rapidly in solution; the racemisation probably occurs *via* enolisation (*cf.* §8 iii II).



In the formation of allantoin from uric acid by oxidation, one carbon atom is lost from the latter as carbon dioxide. The problem, then, is to fit one carbon atom into the allantoin structure. At the same time, the structure thus given to uric acid must also include the alloxan skeleton in order to account for the formation of this compound. Two structures that were proposed which both agreed with the facts known at the time were by Medicus (1875) and by Fittig (1878).



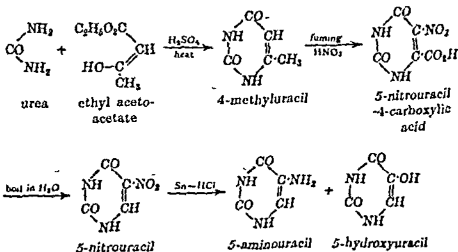
Medicus formula



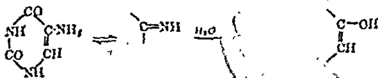
Fittig formula

Fischer (1884) prepared two isomeric monomethyluric acids; one gave methylalloxan and urea on oxidation with nitric acid, and the other gave alloxan and methylurea. Fittig's formula, which is symmetrical, can give rise to only *one* monomethyluric acid; hence this structure is untenable. On the other hand, the Medicus formula satisfies the existence of at least two isomeric monomethyl derivatives: one methyl group in the pyrimidine nucleus (at position 1 or 3) would produce methylalloxan and urea, and a methyl group in the glyoxaline nucleus (at position 7 or 9) would produce alloxan and methylurea (Fischer showed that the two monomethyluric acids were the 3- and 9-derivatives). Examination of the Medicus formula shows that it admits the possibility of four monomethyl, six dimethyl and four trimethyl derivatives. All of these have been prepared by Fischer and his co-workers, thus giving powerful support to the Medicus formula. Proof of the Medicus formula lies in the synthesis of uric acid; three syntheses are given here.

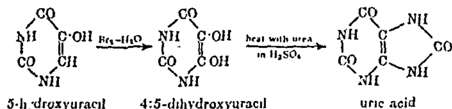
(i) Behrend and Roosen (1888) carried out the first unambiguous synthesis (see also §15. XII).



In this reduction, some of the aminouracil is converted into hydroxyuracil. The mechanism of this change is not certain, but a possibility is as follows:



The reaction product was treated with nitrous acid, thereby converting the 5-aminouracil present into 5-hydroxyuracil; then the synthesis proceeded as follows:



5-hydroxyuracil 4:5-dihydroxyuracil uric acid

(ii) Baeyer's synthesis (1863), completed by Fischer (1895). Bayer arrived at ψ-uric acid and knew that uric acid contained one molecule of water less than this, but was unable to remove it to form uric acid. His failure was due to the fact that ψ-uric acid is not dehydrated by the usual dehydrating agents; Fischer succeeded by fusion with anhydrous oxalic acid, and also obtained better results by boiling ψ-uric acid with 20 per cent. hydrochloric acid.

$CO(NH_2)_2$ + $CH_2(CO_2H)_2$ (malonic acid) $\xrightarrow{POCl_3}$ barbituric acid $\xrightarrow{HNO_3}$ violuric acid ($C{=}NOH$) $\xrightarrow{NH_4HS}$

uramil ($CH·NH_2$) $\xrightarrow[\text{in } H_2O]{KNCO}$ ψ-uric acid ($CH·NH·CO·NH_2$) $\xrightarrow[\text{boil}]{20\%\ HCl}$ [intermediate with $C·OH$ and $CO·NH_2$]

$\xrightarrow{-H_2O}$ uric acid

(iii) Traube's synthesis (1900) is the most important method, since it can be used to prepare any purine derivative; it is also the basis of various commercial methods for preparing the purines synthetically.

$CO(NH_2)_2$ + $C_2H_5O_2C \cdot CH_2 \cdot CN$ $\xrightarrow{POCl_3}$ $NH_2 \cdot CO \cdot NH \cdot CO \cdot CH_2 \cdot CN$ $\xrightarrow{NaOH}$ (cyclic: NH—CO—CH₂—C=NH—NH—CO)

ethyl cyanoacetate; cyanoacetylurea

$\rightleftharpoons$ (CO—CH=C·NH₂—NH—CO—NH ring) $\xrightarrow{HNO_2}$ (CO—C·NO=C·NH₂—NH—CO—NH ring) $\xrightarrow{NH_4HS}$

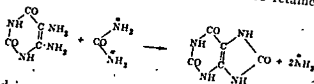
(CO—C·NH₂=C·NH₂—NH—CO—NH ring) $\xrightarrow[NaOH]{Cl\,CO_2C_2H_5}$ (CO—C·NH·CO₂C₂H₅=C·NH₂—NH—CO—NH ring)

4:5-diamino-uracil

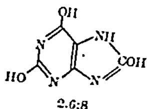
fuse with urea; heat at 180°

(uric acid: CO—NH—C—NH—CO—NH—C—NH—CO—NH)

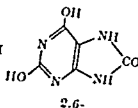
Clusius *et al.* (1953), using urea labelled with ^{15}N, have shown that the two nitrogen atoms in the diaminouracil are retained on fusion with urea.



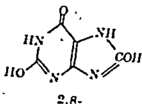
Uric acid is a white crystalline powder which is insoluble in the ordinary organic solvents. It behaves as a weak dibasic acid, forming two series of salts (*e.g.*, monosodium and disodium urate).



2.6:8



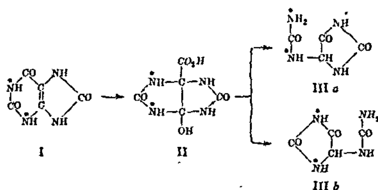
2.6-



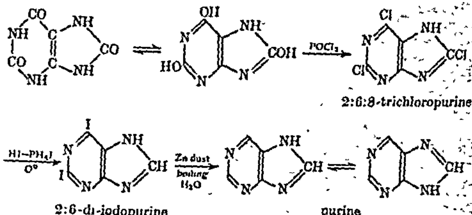
2.8-

It thus appears that the tri-enol form (2 6.8-trihydroxypurine) is unlikely; this leaves three possible di-enol forms, 2.6-, 2:8-, and 6:8-. Which of these di-enol forms is the one that forms the disodium salt still appears to be uncertain. Fischer thought that the di-enol form is the 2:6-. Evidence that may be quoted to support this is that in this arrangement the pyrimidine ring will be "aromatic" and so stabilised by resonance. There is, however, a certain amount of evidence which suggests the 2.8- di-enol form (*cf.* §§ 13a, 15).

It is also interesting to consider the path followed in the oxidation of uric acid to allantoin. Behrend (1904) suggested that the alkaline permanganate oxidation of uric acid (I) gives allantoin (III*a* and *b*) *via* the symmetrical intermediate II. Cavalieri *et al.* (1948) have carried out this oxidation using uric acid labelled with ^{15}N at N_1 and N_3, and found that the allantoin produced had this isotopic nitrogen distributed uniformly among all the four nitrogen atoms. This is in keeping with the intermediate formation of II



§3. **Purine.** When uric acid is treated with phosphoryl chloride, 2.6·8-trichloropurine is obtained (uric acid behaves as the tri-enol in this reaction). This trichloro compound is a very important intermediate in the synthesis of purine derivatives, and a point worth noting is that the reactivities of the chlorine atoms are $6 > 2 > 8$. Purine, m.p. 217°, may be prepared from uric acid as follows:

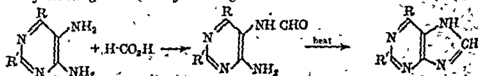


Purine is a fairly strong base and forms salts with acids; it is a synthetic compound.

PURINE DERIVATIVES

§4. Synthesis of purines. Before describing some individual purine derivatives, let us first consider some general methods of synthesising purines. Fischer (1897, 1898) prepared various purines starting from 2:6.8-trichloropurine. There are, however, two general synthetic methods in which the pyrimidine ring is synthesised first and then the glyoxaline ring "built up" on this, or *vice versa*.

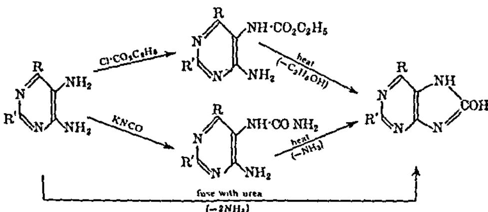
(i) *Traube's method.* This consists of synthesising a 4:5-diaminopyrimidine (see §14. XII) and then condensing with formic acid to produce the glyoxaline ring; the formyl derivative is ring-closed by heating alone or by heating its sodium salt.



This synthesis leads to the preparation of purines that are unsubstituted in position 8. This type of purine may also be prepared by heating a 4:5-diaminopyrimidine with dithioformic acid in the presence of sodium hydroxide solution, and then heating the product with a methanolic solution of sodium methoxide.

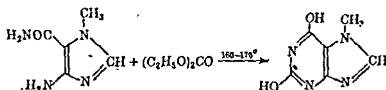


8-Hydroxypurines may be prepared by using ethyl chloroformate instead of formic acid. Alternatively, the diaminopyrimidine may be boiled with potassium *iso*cyanate and the product, a ureidopyrimidine, ring-closed by heating. Finally, diaminopyrimidines may be fused with urea to produce 8-hydroxypurines.



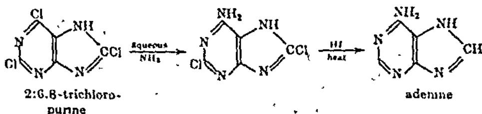
o-Aminohydroxypyrimidines may be used instead of *o*-diaminopyrimidines (*cf.* Baeyer's synthesis of ψ-uric acid, §2).

(ii) A less frequently used synthesis of purines starts with the glyoxaline derivative, *e.g.*, 7-methylxanthine from 4-amino-1-methylglyoxaline-5-carbonamide (Sarasin *et al.*, 1924).

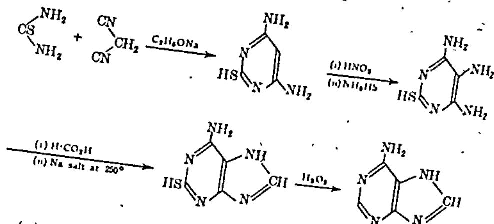


§5. Adenine (6-aminopurine), d. 365°, occurs in the pancreas of cattle and in tea extract. Its general reactions showed that adenine was a purine, and its structure was established by synthesis.

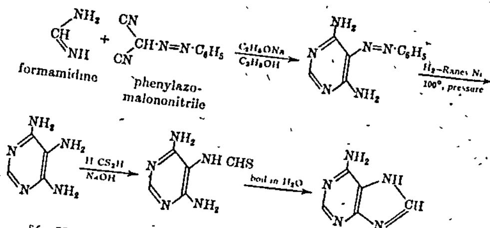
(i) Fischer (1897).



(ii) Traube (1904).

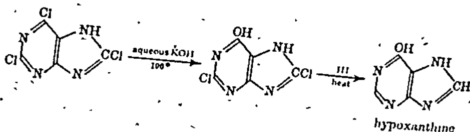


(iii) Todd *et al.* (1943).

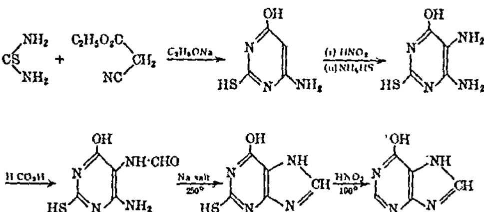


§6. **Hypoxanthine** (6-hydroxypurine), d. 150°, occurs in tea extract and in animal tissues. Its formation by the action of nitrous acid on adenine establishes its structure, and this is confirmed by synthesis.

(i) Fischer (1897, 1898).

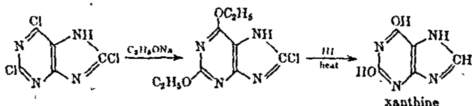


(ii) Traube (1904).

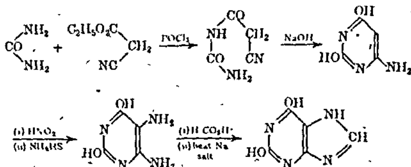


§7. **Xanthine** (2 6-dihydroxypurine), d. above 150°, occurs in tea extract and in animal tissues. When oxidised with potassium chlorate in hydrochloric acid solution, xanthine forms alloxan and urea; these products show the relationship of xanthine to uric acid, and its structure has been established by synthesis.

(i) Fischer (1898).



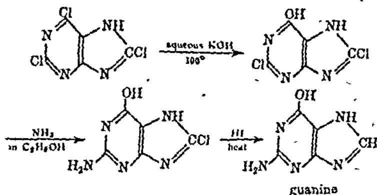
(ii) Traube (1900).



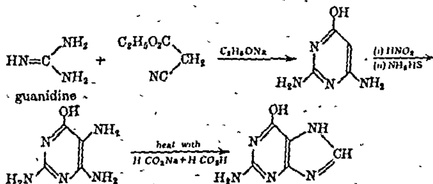
Xanthine is the parent substance of a number of compounds (see later).

§8. Guanine [illegible] 0°, occurs in the [illegible]s. Its structure [illegible] it gives xanthine on treatment with [illegible] acid; this conversion is also effected by boiling guanine with 25 per cent. hydrochloric acid (Fischer, 1910).

(i) Fischer (1897).



(ii) Traube (1900).

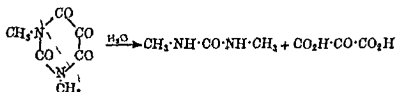


XANTHINE BASES

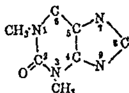
Three important methylated xanthines that occur naturally are caffeine, theobromine and theophylline. All three have been prepared from uric acid by Fischer and all have been synthesised by means of the Traube method.

§9. Caffeine (*1:3:7-trimethylxanthine*), m.p. 235–237°, occurs in tea, coffee, etc. Its molecular formula is $C_8H_{10}O_2N_4$, and its relationship to uric acid is shown by the fact that on oxidation with potassium chlorate in hydrochloric acid, caffeine gives dimethylalloxan and methylurea in equimolecular proportions. The structure of the former product is established by its conversion into

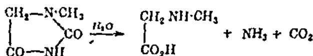
sym.-dimethylurea and mesoxalic acid on hydrolysis, and is confirmed by synthesis from these two compounds.



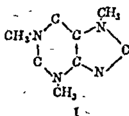
These results indicate that caffeine and uric acid have the same skeleton structure; at the same time the positions of two methyl groups and one oxygen atom in caffeine are also established. Thus the problem now is to ascertain the positions of the remaining methyl group and oxygen atom. The following skeleton structure for caffeine summarises the above information; the third methyl group is at either position 7 or 9, and the remaining oxygen atom at 6 or 8.

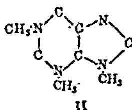


Position of the methyl group. As we have seen above, the oxidation of caffeine gives dimethylalloxan and methylurea. Fischer, however, also isolated another oxidation product which, on hydrolysis, gave *N*-methylglycine, carbon dioxide and ammonia. Thus this third oxidation product must be *N*-methylhydantoin:



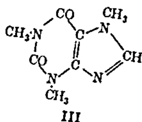
It therefore follows that caffeine contains two ring structures, that of dimethylalloxan and that of methylhydantoin. The following two skeleton structures for caffeine are both possible, since each

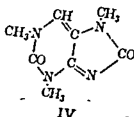




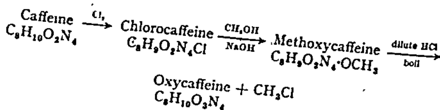
could give the required oxidation products. Actually, the isolation of methylurea suggests I or II; the isolation of methylhydantoin confirms these possibilities. Finally, Fischer isolated a fourth oxidation product, *viz.*, *sym.*-dimethyloxamide, $CH_3\cdot NH\cdot CO\cdot CO\cdot NH\cdot CH_3$. Examination of I and II shows that only I can give rise to the formation of this oxamide, and so I is the skeleton of caffeine.

Position of the oxygen atom. In view of what has been said above, we see that there are now two possible structures for caffeine which fit the facts equally well:



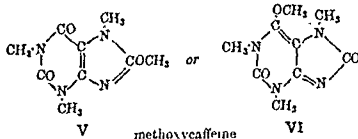


By analogy with uric acid, III would appear the more likely one; this, however, is not proof. Fischer showed that III is caffeine as follows.

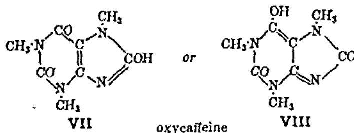


$$\underset{C_8H_{10}O_2N_4}{\text{Caffeine}} \xrightarrow{Cl_2} \underset{C_8H_9O_2N_4Cl}{\text{Chlorocaffeine}} \xrightarrow[NaOH]{CH_3OH} \underset{C_8H_9O_2N_4\cdot OCH_3}{\text{Methoxycaffeine}} \xrightarrow[\text{boil}]{\text{dilute HCl}} \underset{C_8H_{10}O_3N_4}{\text{Oxycaffeine}} + CH_3Cl$$

Fischer then showed that oxycaffeine was identical with a trimethyluric acid, since on methylation with methyl iodide in the presence of aqueous sodium hydroxide, oxycaffeine was converted into tetramethyluric acid. Thus methoxycaffeine is either V or VI, and oxycaffeine VII or VIII.

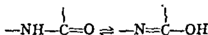


V methoxycaffeine VI



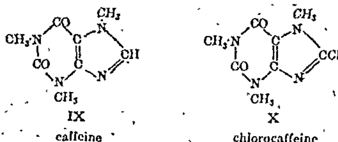
VII oxycaffeine VIII

mation of these two products suggests that oxycaffeine is a tautomeric substance, *i.e.*, it contains the *amido-imidol triad system*:



$$—NH—\overset{|}{C}=O \rightleftharpoons —N=\overset{|}{C}—OH$$

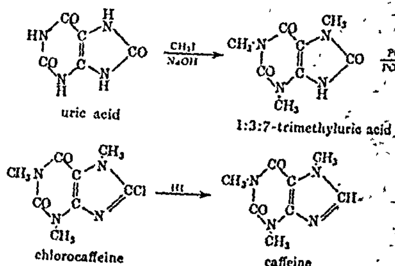
Now this triad system can exist only in the glyoxaline nucleus in oxycaffeine, since neither nitrogen atom in the pyrimidine nucleus is attached to a hydrogen atom (VII can give rise to the above tautomeric system, whereas VIII cannot). Thus the methoxyl group in methoxycaffeine is in the glyoxaline nucleus, and consequently the chlorine atom in chlorocaffeine is also in this nucleus; hence caffeine is IX and chlorocaffeine is X.



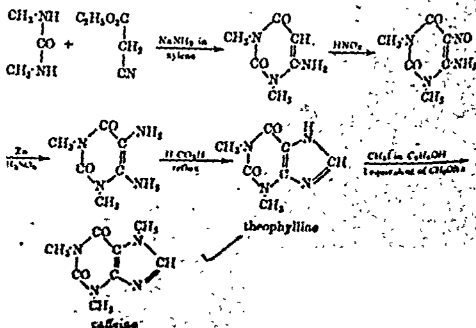
IX caffeine X chlorocaffeine

This structure for caffeine has been confirmed by various syntheses, *e.g.*,

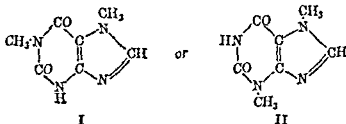
(i) Fischer (1899).



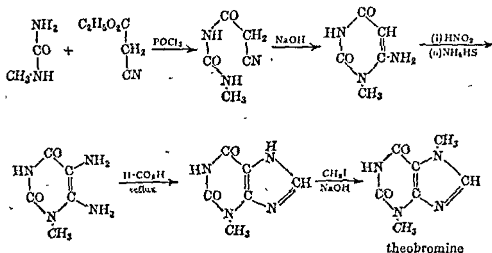
(ii) A commercial synthesis based on Traube's method is as follows:



§10. Theobromine (3:7-dimethylxanthine), m.p. 337°, occurs in cocoa beans, tea, etc. The structure of theobromine has been deduced from the fact that, on oxidation with potassium chlorate in hydrochloric acid, it gives methylalloxan and methylurea, and also that it is converted into caffeine when its silver salt is heated with methyl iodide. Thus theobromine is either I or II.

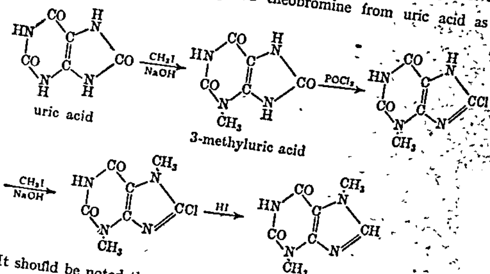


The position of the methyl group in the pyrimidine nucleus has been shown to be 3 (*i.e.*, structure II) by synthesis using Traube's method.



The product formed by the condensation between methylurea and ethyl cyanoacetate contained no free amino group; thus the condensation must occur as shown (and not by the carbethoxyl group with the methylimino group of the methylurea).

Fischer (1899) also prepared theobromine from uric acid as follows.



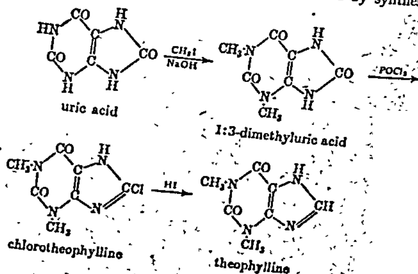
It should be noted that in this synthesis a mixture of phosphorus pentachloride and phosphoryl chloride cannot be used; this mixture replaces the oxygen atom (*i.e.*, the hydroxyl group) at position 6 and not at 8.

§11. Theophylline (1:3-dimethylxanthine) ... in tea. Its structure ...

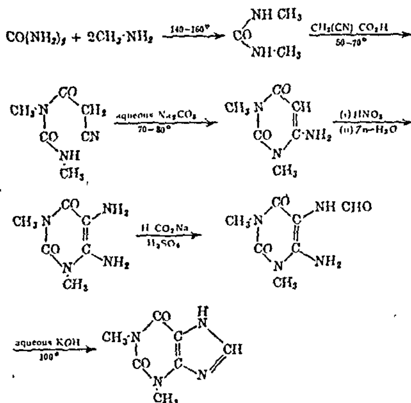
... this structure has been ... theophylline is 1:3-dimethyl- ...

(i) Fischer (1899).

... structure has been confirmed by synthesis.



(ii) Theophylline has also been synthesised commercially by means of the Traube method (*cf.* caffeine, §9).

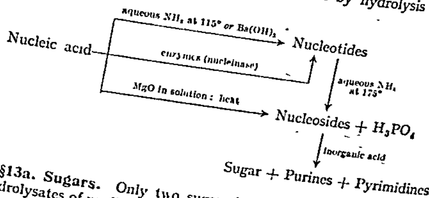


§11a. Biosynthesis of purines. Most of the work on the biosynthesis of purines has been carried out on uric acid by means of enzymes from bird liver. Sonne *et al.* (1946), working with the following labelled compounds (^{14}C), showed that carbon dioxide supplies C_6, formic acid C_2 and C_8, and glycine C_4, C_5 and N_7. Thus all the carbon atoms in uric acid are accounted for, and the three remaining nitrogen atoms, N_1, N_3 and N_9, are believed to be derived from ammonia (provided by the metabolism of amino-acids). The actual sequence of the steps involved in purine synthesis is unknown, but there is some evidence to show that the various methylated xanthines are produced by progressive methylation. It has been shown, however, that xanthine is not an intermediate in the biosynthesis of uric acid, and hence it is possible that uric acid and the xanthine bases are synthesised by different routes.

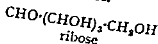
NUCLEIC ACIDS

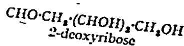
§12. Introduction. Nucleoproteins are one of the classes of conjugated proteins (§7 B. XIII); the nucleic acid part is the prosthetic group, and the protein part consists of protamines and histones. These latter compounds are basic and form salt-like compounds, the nucleoproteins, with the nucleic acids. On careful hydrolysis, nucleoproteins are broken down into the nucleic acid and protein.

§13. Structure of the nucleic acids. Nucleic acids are colourless solids, all of which contain the following elements: carbon, hydrogen, oxygen, nitrogen and phosphorus. The following chart shows the nature of the products obtained by hydrolysis under different conditions.



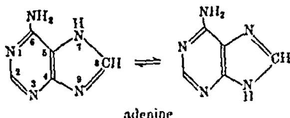
§13a. Sugars. Only two sugars have been isolated from the hydrolysates of nucleic acids; both are pentoses: D(−)-ribose and 2-deoxy-D(−)-ribose.



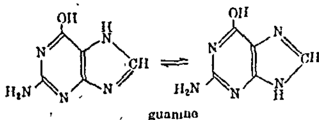


The nucleic acids are classified according to the nature of the sugar present: the *pentose nucleic acids* or *ribonucleic acids* (R.N.A.), and the *deoxypentose nucleic acids* or *deoxyribonucleic acids* (D.N.A.). Ribonucleoproteins are found mainly in the cytoplasm of the cell, whereas deoxyribonucleoproteins are found mainly in the cell nucleus. D(−)-Ribose is the pentose of yeast, liver and pancreas R.N.A.s; 2-deoxy-D(−)-ribose occurs in thymus D.N.A.

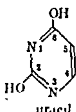
§13b. Bases. Only two purine bases have been obtained from nucleic acids, adenine and guanine. On the other hand, five



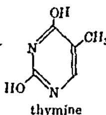
adenine



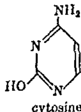
guanine



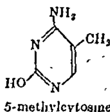
uracil



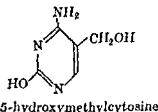
thymine



cytosine



5-methylcytosine



5-hydroxymethylcytosine

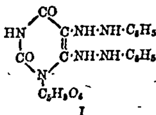
pyrimidine bases have been isolated: uracil, thymine, cytosine, 5-methylcytosine and 5-hydroxymethylcytosine. Both types of nucleic acids (R.N.A.s and D.N.A.s) contain adenine and guanine. Cytosine also occurs in both types of nucleic acids, but uracil occurs only in R.N.A.s, and thymine only in D.N.A.s. 5-Methylcytosine has been found to be a fairly common minor constituent of D.N.A.s; it has not yet been found in R.N.A.s (Wyatt, 1950). 5-Hydroxymethylcytosine has been found in certain D.N.A.s (Wyatt *et al.*, 1952).

Combination of a base (either a purine or pyrimidine) with a sugar (ribose or deoxyribose) gives rise to a **nucleoside**, *e.g.*, adenosine (ribose + adenine), guanosine (ribose + guanine), cytidine (ribose + cytosine), uridine (ribose + uracil), thymidine (deoxyribose + thymine).

Combination of a nucleoside with phosphoric acid produces a **nucleotide**, *i.e.*, nucleotides are nucleoside phosphates, *e.g.*, adenylic, guanylic, cytidylic and uridylic acids. It might be noted here that the term nucleotide is now used to embrace a large group of compounds composed of the phosphates of *N*-glycosides of heterocyclic bases, and the pyrophosphates and polyphosphates containing one or more nucleosides. The term nucleotide also includes the nucleic acids themselves.

The problem now is to ascertain how these various units are linked in nucleosides and nucleotides.

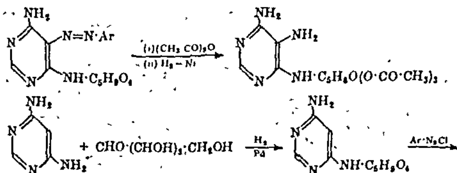
§13c. Structure of nucleosides. Hydrolysis of nucleotides with aqueous ammonia at 175° under pressure gives nucleosides and phosphoric acid; thus in nucleosides the base is linked directly to the sugar. Furthermore, since nucleosides are non-reducing, the "aldehyde group" of the sugar cannot be free, *i.e.*, nucleosides are *glycosides* (*cf.* §24. VII). The next problem is to decide which atom of the base is joined to C_1 of the sugar. Let us first consider the pyrimidines. Cytidine, on treatment with nitrous acid, is converted into uridine; it therefore follows that the sugar residue is linked in the *same* position in both of these nucleosides. The point of linkage cannot be 1 or 6, since cytidine has a *free* amino group at position 6 and consequently there cannot be a hydrogen atom on N_1. Also, since uridine forms a 5-bromo derivative, C_5 must be free (Levene *et al.*, 1912). When uridine is treated with an excess of bromine, followed by the addition of phenylhydrazine, a uridine derivative is obtained which contains *two* phenylhydrazino radicals. This compound was given structure I since work by Levene (1925) showed that this type of compound can be obtained



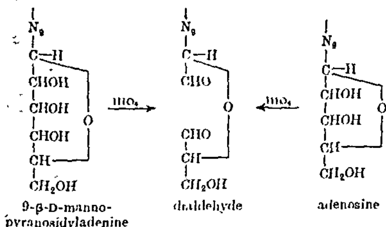
only if uracil is substituted in position 3 and positions 4 and 5 are free. Thus the sugar is attached to N_3. In a similar way, it has been shown that the other pyrimidine nucleosides (ribosides and deoxyribosides) have the sugar residue linked at N_3. Todd *et al.* (1947) have synthesised uridine and cytidine, and thereby

have confirmed the linkage at N_3. This linkage has also been confirmed by the X-ray analysis of cytidine (Furberg, 1950).

Now let us consider nucleosides containing purine bases. Adenosine has a free amino group at position 6; therefore the sugar cannot be at C_6 or N_1 (*cf.* cytidine). Similarly, since guanosine has a free amino group at position 2, the sugar cannot be at C_2 or N_3. Now Levene found that the two purine ribosides are equally readily hydrolysed by dilute acids and by the same enzyme. He therefore assumed that the sugar residue is linked at the same place in both nucleosides. On this basis, only positions 7, 8 and 9 are possible points of attachment. Position 8 was then excluded since this point would involve a carbon-carbon bond, a linkage which would be very stable, whereas nucleosides are very readily hydrolysed by dilute acids (see also below). Thus positions 7 or 9 are free. This is supported by the following evidence (Levene, 1923). When guanosine is treated with nitrous acid, xanthosine is produced and this, on methylation with diazomethane followed by hydrolysis, gives theophylline (1:3-dimethylxanthine). Thus positions 1 and 3 are free in guanosine, and so the sugar must be attached at position 7 or 9. The evidence so far does not permit a decision to be made between these two positions since the system (in the glyoxaline nucleus) is tautomeric. It should be noted that had the sugar residue been attached to C_8, then a *trimethylxanthine* would have been obtained instead of theophylline (*cf.* above). The ultraviolet absorption spectrum of guanosine is very similar to that of 9-methylguanine and differs from that of 7-methylguanine; hence it appears likely that guanosine is the 9-guanine glycoside (Gulland *et al.*, 1936, 1938). Todd *et al.* (1947, 1948) have synthesised guanosine and adenosine in which the sugar is known to be in the 9-position, and showed that their synthetic compounds are identical with the natural products; *e.g.*, the synthesis of adenosine.



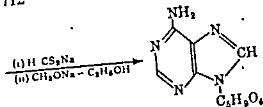
The configuration of the furanoside link has been shown to be β- by various means, *e g.*, Todd *et al.* (1947) oxidised adenosine with periodic acid, and showed that the product is identical with that from the oxidation of 9-β-D-mannopyranosidyladenine (a syn-



thetic compound). This proves that the sugar residue is at position 9, has the furanose structure, and that the linkage is β-. Similar

II + III →

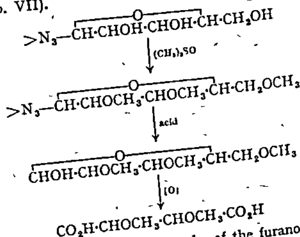
$\xrightarrow{NH_3 \text{ in } CH_3OH}$ IV $\xrightarrow{H_2-Pd}$ adenosine



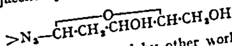
adenosine

It might be noted, in passing, that glycosides are compounds formed by the linking of a sugar (at C_1) with a COH group. Thus the nucleosides are, strictly speaking, not glycosides; they should be called ribosyl-pyrimidines and ribosyl-purines.

The final problem to be elucidated in connection with the structure of nucleosides is the nature of the ring in the sugar residue and the type of linkage (α or β). Degradative experiments have shown that the sugar is present as the furanose form, *e.g.*, methylation of a pyrimidine riboside, followed by hydrolysis, gives a trimethylribose which, on oxidation, forms dimethyl*meso*tartaric acid. This product shows that the ribose ring is furanose; had the ring been pyranose, then the final product would have been trimethoxyglutaric acid (*cf.* §§7a, 7b. VII).

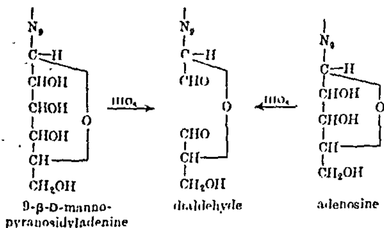


Deoxyribose has also been shown to be of the furanose type, *e.g.*, Lythgoe *et al.* (1950) found that pyrimidine deoxyribosides consume a negligible amount of periodic acid; this agrees with the 2-deoxyribofuranose structure since, in this state, the molecule does not contain two adjacent hydroxyl groups (*cf.* §7g. VII).



These results have been confirmed by other work (see below).

The configuration of the furanoside link has been shown to be β- by various means, *e.g.*, Todd *et al.* (1947) oxidised adenosine with periodic acid, and showed that the product is identical with that from the oxidation of 9-β-D-mannopyranosidyladenine (a syn-



thetic compound). This proves that the sugar residue is at position 9, has the furanose structure, and that the linkage is β-. Similar

II

III

NH_3 in CH_3OH

H_2–Pd

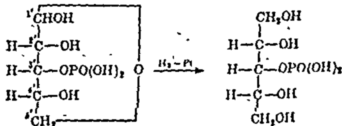
IV

adenosine

experiments with other ribonucleosides suggest that all these compounds have a β-configuration. Also, Todd *et al.* (1946–1948) have synthesised adenosine, guanosine, cytidine and uridine, and thereby confirmed the β-configuration; *e.g.*, adenosine has been synthesised as follows (Todd *et al.*, 1948). Acetochloro-D-ribofuranose, II (*cf.* §24. VII), is condensed with the silver salt of 2:8-dichloroadenine, III, and the product deacetylated with a methanolic solution of ammonia to give 2:8-dichloro-9-β-ribofuranosyladenine, IV. IV, on catalytic reduction (palladium), is converted into adenosine.

Furberg (1950) has shown by means of the X-ray analysis of cytidine that the sugar residue is attached to N_3 and is β-D-ribofuranoside. Since other ribonucleosides exhibit the same general pattern, it is inferred that all are furanosides with the β-configuration. Manson *et al.* (1951), from absorption spectra measurements, have shown that deoxyribonucleosides also exist in the β-configuration.

§13d. Structure of nucleotides. When nucleotides are carefully hydrolysed, ribose monophosphate may be isolated from the products; thus the phosphoric acid is attached to the sugar residue in nucleotides. Examination of the nucleoside structures shows that the point of attachment may be 2′, 3′ or 5′ in the ribose molecule, and 3′ or 5′ in the deoxyribose molecule. On reduction with hydrogen in the presence of platinum, ribose phosphate is converted into an optically inactive phosphoribitol (Levene *et al.*, 1932, 1933). This product can be optically inactive only if the phosphate residue is attached to the *centre* hydroxyl group of the ribose molecule, *i.e.*, at the 3′-position.



It should be remembered that the furanose structure occurs only when the sugar is in the form of a glycoside; on hydrolysis, the furanose sugar first liberated immediately changes into the stable pyranose form (see §7*f*. VII).

Until recently, it was believed that the 3′-position was the only one occupied by the phosphate radical. Emden *et al.* (1929) claimed to have isolated a 5′-phosphate (from muscle nucleic acid). Carter and Cohn (1949) isolated two isomeric adenylic acids from

the alkaline hydrolysates of R.N.A.s, and called them "*a*" and "*b*" adenylic acids. These authors, in 1950, also isolated two isomers of guanylic, uridylic and cytidylic acids. Carter and Cohn found that one of their adenylic acids was identical with adenosine-3′ phosphate, but the other was not the same as the 5′-compound of Emden. These authors therefore believed that their two isomers were the 2′- and 3′-phosphate. Todd *et al.* (1952) synthesised adenosine-2′ and 3′ phosphate, and showed that their synthetic compounds were identical with the "*a*" and "*b*" acids obtained by Carter and Cohn, but were not able to say which was which. Loring *et al.* (1952) showed that the "*a*" and "*b*" cytidylic acids resist oxidation by periodic acid, and hence it follows that they must be the 2′- and 3′-phosphates (but there is no indication from this which isomer is the 2′- and which is the 3′-); had one isomer been the 5′-compound, then it would have been oxidised by periodic acid (the two hydroxyls on 2′ and 3′ are free and adjacent). A study of the solubility, acidity and absorption spectra of these two cytidylic acids led Loring *et al.* to suggest that the "*a*" acid is the 2′-phosphate. This conclusion has been supported by Harris *et al.* (1953) from their study of the infra-red spectra of these compounds. Todd *et al.* (1954) have synthesised deoxycytidine-3′ phosphate, and comparison of its infra-red spectra and other properties with cytidine phosphates provides strong evidence that "*b*" cytidylic acid is cytidine-3′ phosphate, and therefore that "*b*" uridylic acid is uridine-3′ phosphate. Brown *et al.* (1955) have shown that hydrazine splits "*a*" and "*b*" cytidylic acid to give ribose 2- and 3-phosphate respectively. "*b*" Uridylic acid yields the same ribose phosphate obtained from "*b*" cytidylic acid. Thus the "*a*" and "*b*" isomers of these nucleotides are the 2′- and 3′-phosphates, respectively, of the ribonucleosides.

Experiments using enzymic hydrolysis of nucleic acids have shown that these acids also contain 5′-phosphoester links. Cohn *et al.* (1951) have isolated the 5′-phosphates of adenosine, guanosine, uridine and cytidine. These authors have also shown that the nucleotides in calf thymus D.N.A. are 3′- and 5′-phosphates (position 2′ is not possible since this is a CH_2 group).

Thus, according to the foregoing evidence, the phosphate radical can occupy the positions 2′, 3′ and 5′ in ribonucleotides, and 3′ and 5′ in deoxyribonucleotides. These, however, by no means exhaust the possible positions of the phosphate radical. Todd *et al.* (1951) have identified *cyclic* nucleoside phosphates (2′:3′-) from the enzymic hydrolysates of R.N.A.s. If these cyclic esters are actually present in nucleic acids, then the 2′- and 3′-phosphates obtained by hydrolysis may arise by the opening of the cyclic compound

This simple structure for nucleic acids has, however, been shown to be incorrect by more recent work, *e.g.*,

(i) It has been found that alkaline methods of purification degrade nucleic acids; thus the molecular weight varies with the methods used for the isolation of the acid.

(ii) Various methods for determining molecular weights, *e.g.*, diffusion and the ultracentrifuge, have shown that the D.N.A. from the thymus gland has a molecular weight of 500,000 to 1,000,000 or more; R.N.A. from yeast has a molecular weight of 6,500 to 290,000.

(iii) Various investigations, *e.g.*, X-ray studies and light scattering effects, indicate that D.N.A. molecules have little, if any, branching. On the other hand, R.N.A. molecules appear to have a branched structure. X-ray studies have shown that D.N.A.'s are composed of two polynucleotide chains wound as spirals round a common axis but head in opposite directions (Wilkins *et al.*, 1953; Watson *et al.*, 1953). Furthermore, in the solid state, the helix may be either compressed or somewhat extended.

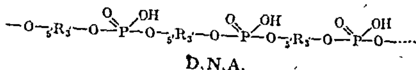
(iv) The analysis of the hydrolysates of nucleic acids, particularly by chromatography, has shown that the acids from different sources have different chemical compositions. According to Chargaff (1950), not one specimen of a nucleic acid gave analysis results corresponding to a tetranucleotide; thus the "statistical tetranucleotide" theory is untenable. Chargaff found that in D.N.A.s, the sum of the total purine nucleotides is equal to that of the pyrimidine nucleotides, and that the molar ratios of adenine to thymine, and of guanine to cytosine (or its analogues) are unity. Chargaff *et al.* (1954) also found the same regularities in R.N.A.s, with uracil taking the place of thymine. Chargaff estimated the nucleotide content from spectral data (as well as by some of the earlier methods), and pointed out that the regularities are not usually observed with *purified* samples of pentose nucleic acids, but only when, *e.g.*, whole cells are subjected to hydrolysis.

(v) Levene *et al.* (1926), from electromeric titration experiments, concluded that R.N.A.s show four primary and one secondary phosphate dissociation for each set of four phosphorus atoms present. On this evidence, and on the results of analysis, Levene put forward his tetranucleotide theory (see above). More refined titration experiments, however, have shown that R.N.A.'s exhibit only three primary and one secondary phosphate dissociation (Gulland *et al.*, 1944). These latter findings are also supported by methylation experiments (Anderson *et al.*, 1949).

(vi) The sequence of nucleotides in nucleic acids is not yet known; the number of permutations is enormous (*cf.* proteins, §8. XIII).

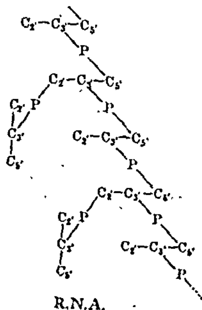
The evidence obtained so far appears to indicate that yeast ribonucleic acid consists of chains of purine nucleotides linked to chains of pyrimidine nucleotides.

Various structures have been proposed for the nucleic acids, *e.g.*, Todd (1952) has suggested the following for deoxyribonucleic acids:



D.N.A.

Todd has also suggested the following for ribonucleic acids:



R.N.A.

Todd *et al.* (1953) have carried out some preliminary experiments which may lead to the solution of the sequence of nucleotides in nucleic acids. Jones *et al.* (1956) have also developed a chemical method for the specific degradation of deoxyribonucleic acids. These authors have found that on treatment with mercaptoacetic acid ($CH_2SH{\cdot}CO_2H$), purines are removed and replaced by carboxymethylthio groups. By this means it is possible to obtain information on the relative positions of purines and pyrimidines. Thus the results have shown that in calf-thymus deoxyribonucleic acids there are regions in which at least three pyrimidine nucleotides occur in adjacent positions.

READING REFERENCES

Fischer, Synthesen in der Puringruppe, *Ber*, 1899, **32**, 435.
Stewart, *Recent Advances in Organic Chemistry*, Longmans, Green Vol. I (1931, 6th ed) Ch 13. The Purine Group.
Lythgoe, Some Aspects of Pyrimidine and Purine Chemistry, *Quart Reviews* (*Chem. Soc.*), 1949, **3**, 181
Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1938, 1st ed) Vol. II. Ch 11. The Chemistry of Pyrimidines, Purines and Nucleic Acids.
Levene and Bass, *Nucleic Acids*, Chemical Catalogue Co (1931)
Davidson, *The Biochemistry of the Nucleic Acids*, Methuen (1953, 2nd ed).
Ann. Reports (*Chem Soc.*). (i) 1944, **41**, 200 Chemistry of Nucleosides and Nucleotides. (ii) 1950, **47**, 253; 1952, **49**, 217, 246 Nucleic Acids, Nucleosides, and Nucleotides (iii) 1954, **51**, 270 Polynucleotides. (iv) 1955, **52**, 393, 1956, **53**, 260 Nucleic Acids.
Tipson, The Chemistry of the Nucleic Acids, *Advances in Carbohydrate Chemistry*, Academic Press, Vol. I (1945)
Schlenk, Chemistry and Enzymology of Nucleic Acids, *Advances in Enzymology*, Interscience Publishers, 1949, **9**, 455.
Gulland, Some Aspects of the Chemistry of Nucleotides, *J C.S.*, **1944**, 208
Todd, Synthesis in the Study of Nucleotides, *J C.S*, **1946**, 647
Markham *et al*, Evidence for the Existence of Two Types of Ribonucleic Acid, *Nature*, 1954, **173**, 537.
Todd, Nucleic Acid and Function, *Chem. and Ind*, **1955**, 1139.
Avison and Hawkins, The Role of Phosphoric Esters in Biological Reactions, *Quart. Reviews* (*Chem. Soc*), 1951, **5**, 171.

CHAPTER XVII

VITAMINS

§1. Introduction. In addition to oxygen, water, proteins, fats, carbohydrates and certain inorganic salts, a number of organic compounds are also necessary for the life, growth and health of animals (including man). These compounds are known as the "accessory dietary factors" or *vitamins*, and are only necessary in very small amounts.

Many vitamins have now been isolated and their structures elucidated. As each vitamin was isolated, it was named by a letter of the alphabet, but once its structure had been established (or almost established), the vitamin has generally been renamed (see text).

The vitamins have been arbitrarily classified into the "fat-soluble group" (vitamins A, D, E and K), and the "water-soluble group" (the remainder of the vitamins).

A number of vitamins have already been dealt with in various chapters dealing with natural products with which these particular vitamins are closely associated chemically, *viz.*, vitamins A_1 and A_2 (§7. IX), vitamin C (§11. VII), and the vitamin D group (§§6, 6a, 6b. XI). This chapter is devoted to a number of other vitamins (see the reading references for further information).

From the point of view of chemical structure, there is very little common to the various vitamins, but from the point of view of chemical reactions, many of the water-soluble vitamins have one feature in common, and that is their ability to take part in reversible oxidation-reduction processes. Thus they form a part of various [illegible] VIII), *e.g.*, nicotinamide is present in co- [illegible] DPN), and in co-enzyme [illegible] N); phosphorylated pyridoxal is the co-enzyme of trans[illegible], riboflavin in flavin adenine nucleotide (FAD); pantothenic acid in co-enzyme A; etc.

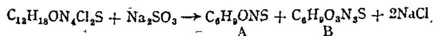
VITAMIN B COMPLEX

§2. Introduction. Eijkman (1897) found that birds developed polyneuritis when fed with polished rice, and were cured when

they were given rice polishings. Then Grijns (1901) found that rice polishings cured beriberi in man (beriberi in man corresponds to polyneuritis in birds; it is a form of paralysis). Grijns suggested that the cause of this paralysis was due to some "deficiency" in the diet, and this was confirmed by Funk (1911, 1912), who prepared a concentrate of the active substance from rice polishings. Funk believed that this active substance was a definite chemical compound, and since he separated organic bases when he prepared his concentrate, he named his "deficiency compound" a *vitamine*. It was then found that "vitamine B" was a complex mixture, and when a number of "vitamines" were obtained that contained no nitrogen, the name *vitamin* was retained for them. The name vitamin B is now reserved for the complex mixture of vitamins in this group.

§3. Vitamin B_1, thiamine (aneurin). Thiamine is one member of the water-soluble vitamin B complex, and is in the thermolabile fraction; it is the absence of thiamine which is the cause of beriberi in man; thus this vitamin is the antineuritic factor (hence the name *aneurin*). Rice polishings and yeast have been the usual sources of thiamine; eggs are also a rich source.

Thiamine is obtained crystalline in the form of its salts; the chloride hydrochloride has been shown to have the molecular formula $C_{12}H_{18}ON_4Cl_2S$ (Windaus *et al.*, 1932); this salt is isolated in the form of its hemihydrate, d. 248–250°. When treated with a sodium sulphite solution saturated with sulphur dioxide at room temperature, thiamine is decomposed quantitatively into two compounds which, for convenience, we shall label A and B (R. R. Williams *et al.*, 1935).

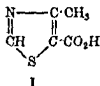


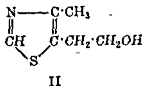
$$C_{12}H_{18}ON_4Cl_2S + Na_2SO_3 \longrightarrow \underset{A}{C_6H_9ONS} + \underset{B}{C_6H_9O_3N_3S} + 2NaCl$$

Compound A, C_6H_9ONS. This compound shows basic properties, and since it does not react with nitrous acid, it was inferred that the nitrogen atom is in the tertiary state. The functional nature of the oxygen atom was shown to be alcoholic, *e.g.*, when A is treated with hydrochloric acid, a hydroxyl group (one oxygen atom and one hydrogen atom) is replaced by a chlorine atom. Furthermore, since the absorption spectrum of the chloro derivative is almost the same as that of the parent (hydroxy) compound, this suggests that the hydroxyl group is in a side-chain. The sulphur did not give the reactions of a mercapto compound nor of a sulphide; in fact, the stability (*i.e.*, unreactivity) of this sulphur atom led to the suggestion that it was in a heterocyclic ring. This conclusion

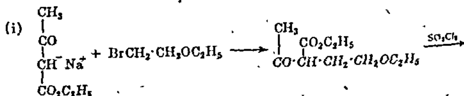
was confirmed by the fact that A has an absorption spectrum characteristic of a thiazole (§5. XII).

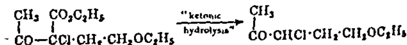
R. R. Williams *et al.* (1935) found that oxidation of A with nitric acid gives the compound $C_5H_5O_2NS$, which can also be obtained by the direct oxidation of thiamine with nitric acid. This latter reaction had actually been carried out by Windaus *et al.* (1934), but these workers had not recognised the presence of the *thiazole* nucleus. Williams *et al.* showed that this oxidation product was a monocarboxylic acid, and found that it was identical with 4-methylthiazole-5-carboxylic acid, I, a compound already described in the literature (Wohmann, 1890). From this it follows that A has a side-chain of two carbon atoms in place of the carboxyl group in I

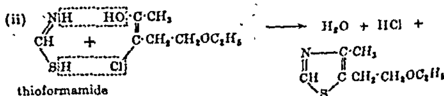


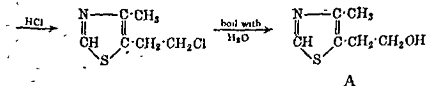


(one carbon atom is lost when A is oxidised to I). Since it is this side-chain which must contain the alcoholic group, the *side-chain* could be either $—CH_2·CH_2OH$ or $—CHOH·CH_3$. Either of these could lose a carbon atom to form a carboxyl group directly attached to the thiazole nucleus. The second alternative, $—CHOH·CH_3$, was excluded by the fact that A does not give the iodoform test, and that A is not optically active (the second alternative contains an asymmetric carbon atom). Thus A was given structure II, and this has been confirmed by synthesis (Clarke *et al.*, 1935).



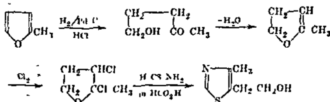




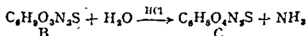


The hydrochloride of this compound is identical with that of the product obtained from thiamine (by fission), and also gives I on oxidation with nitric acid.

Londergan *et al.* (1953) have synthesised A from 2-methylfuran as follows:



Compound B, $C_6H_9O_3N_3S$. This was shown to be a sulphonic acid, *e.g.*, when heated with water under pressure at 200°, B gives sulphuric acid; it also forms sodium sulphite when heated with concentrated sodium hydroxide solution. On treatment with nitrous acid, B evolves nitrogen; thus B contains one or more amino groups. Analysis of the product showed that one amino group is present in B (the product contained only one hydroxyl group). Furthermore, since the evolution of nitrogen was slow, and the reaction of B with benzoyl chloride was also slow, this suggests that B contains an amidine structure (Williams *et al.*, 1935). Williams *et al.* (1935) then heated B with hydrochloric acid at 150° under pressure, and obtained compound C and ammonia. The

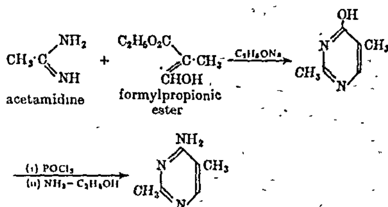


$$\underset{B}{C_6H_9O_3N_3S} + H_2O \xrightarrow{HCl} \underset{C}{C_6H_8O_4N_2S} + NH_3$$

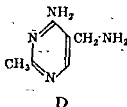
formation of ammonia indicates the replacement of an amino group by a hydroxyl group. This type of reaction is characteristic of 2- and 6-aminopyrimidines; it was therefore inferred that B is a pyrimidine derivative (cf. §14. XII). This is supported by the fact that the ultraviolet absorption spectrum of compound C was similar to that of synthetic 6-hydroxypyrimidines; thus B is probably a 6-aminopyrimidine.

When B is reduced with sodium in liquid ammonia, a sulphonic acid group is eliminated with the formation of an aminodimethylpyrimidine (Williams, 1936). Comparison of the ultraviolet absorption spectrum of this product with various synthetic compounds

showed that it was 6-amino-2:5-dimethylpyrimidine, and this was confirmed by synthesis (Williams *et al.*, 1937).



Thus B is 6-amino-2:5-dimethylpyrimidine with one hydrogen atom (other than one of the amino group) replaced by a sulphonic acid group. When thiamine is treated with sodium in liquid ammonia, one of the products is the diamino derivative D, $C_6H_{10}N_4$ (Williams *et al.*, 1937). Compound D was identified as 6-amino-5-aminomethyl-2-methylpyrimidine by comparison with the absorption

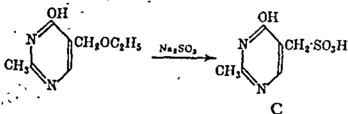


D

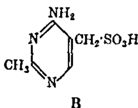
spectra of methylated aminopyrimidines of known structure (Williams *et al.*, 1937). This is confirmed by the synthesis of Grewe (1936). Williams *et al.* had arrived at their conclusion independently of ([illegible]

D, [illegible]

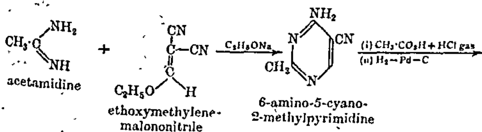
Williams therefore concluded that the sulph[illegible] is joined to the methyl group at position 5. This was confirmed (in 1937) by treating 5-ethoxymethyl-6-hydroxy-2-methylpyrimidine (see the synthesis described for thiamine) with sodium sulphite, whereby 6-hydroxy-2-methylpyrimidyl-5-methanesulphonic acid was obtained, and this was shown to be identical with compound C.

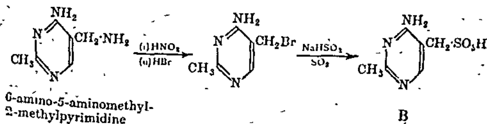


Thus B has the following structure:



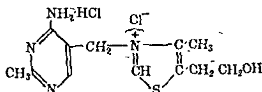
This structure is confirmed by synthesis (Grewe, 1936, Andersag *et al.*, 1937).





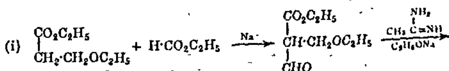
The final problem is: How are fragments A and B united in thiamine? As we have seen, the sulphonic acid group in B is introduced during the fission of thiamine with sodium sulphite; thus the point of attachment of fragment B is at the CH_2 group at position 5. To account for the formation of compound D,

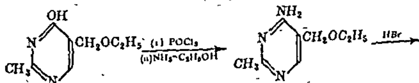
fragment B must be linked to the nitrogen atom of fragment A in this position, the nitrogen atom of the thiazole ring is in a quaternary state, and so accounts for the chloride hydrochloride of thiamine. Had B been connected to A through a carbon atom of the latter, it would not be easy to account for the ready fission of this carbon-carbon bond by means of sodium and liquid ammonia, nor for the fact that thiamine does not form a *dihydrochloride*. Thus the chloride hydrochloride of thiamine is

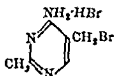


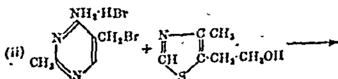
thiamine chloride hydrochloride

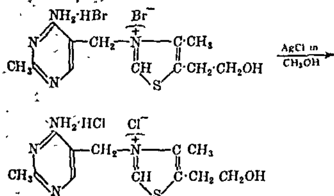
This structure has been confirmed by synthesis, *e.g.*, that of Williams *et al.* (1936, 1937).

(i)


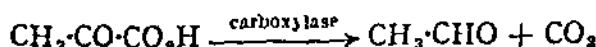




(ii)


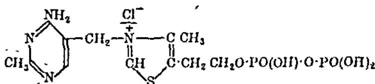


§4. Co-carboxylase. This is the co-enzyme of *carboxylase*, and has been shown to be the pyrophosphate of thiamine (Lohmann *et al.*, 1937). Carboxylase, which requires the co-enzyme for action (see §15. XIII), breaks down pyruvic acid, formed in alcoholic fermentation, to acetaldehyde and carbon dioxide.

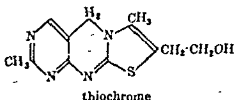


$$CH_3{\cdot}CO{\cdot}CO_2H \xrightarrow{\text{carboxylase}} CH_3{\cdot}CHO + CO_2$$

Co-carboxylase is



§5. Thiochrome was isolated from yeast by Kuhn *et al.* (1935); it is a yellow basic solid and its solutions show a blue fluorescence. Thiochrome is also formed by the oxidation of thiamine with alkaline potassium ferricyanide (Todd *et al.*, 1935); it has also been synthesised by Todd *et al.* (1936).



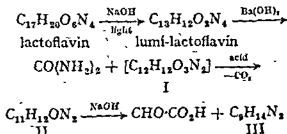
thiochrome

§6. Vitamin B_2, riboflavin (lactoflavin), $C_{17}H_{20}O_6N_4$. Riboflavin is a water-soluble, thermostable vitamin which occurs in the

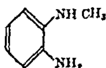
vitamin B complex. It is necessary for growth and health, and occurs widely distributed in nature, *e.g.*, in yeast, green vegetables, milk, meat, etc. Chemically, vitamin B_2 is closely related to the yellow water-soluble pigments known as the *flavins* (*iso*alloxazines), and since it was first isolated from milk, vitamin B_2 is also known as *lactoflavin*.

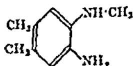
Riboflavin is a bright yellow powder, m.p. 292°, showing a green fluorescence; it is soluble in water and in ethanol, but is insoluble in chloroform and other organic solvents.

When exposed to light, lactoflavin in sodium hydroxide solution forms mainly lumi-lactoflavin, $C_{13}H_{12}O_2N_4$ (this is soluble in chloroform). Lumi-lactoflavin, on boiling with barium hydroxide solution, is hydrolysed to one molecule of urea and one molecule of the barium salt of a β-ketocarboxylic acid, I, $C_{12}H_{12}O_3N_2$ (Kuhn *et al.*, 1933, 1934). The nature of this acid is shown by the fact that, on acidification of the barium salt, the free acid immediately eliminates carbon dioxide to form the compound, II, $C_{11}H_{12}ON_2$. This compound showed the properties of a lactam, and on vigorous hydrolysis by boiling with sodium hydroxide solution, it forms one molecule of glyoxylic acid and one molecule of the compound $C_9H_{14}N_2$ (III).



The structure of III was elucidated as follows (Kuhn *et al.*, 1934). Preliminary tests showed that III was an aromatic diamino compound. Then it was found that it gave a blue precipitate with ferric chloride, and since this reaction is characteristic of *mono*-methyl-*o*-phenylenediamine, it suggests that III contains the following nucleus, IV. The molecular formula of IV is $C_7H_{10}N_2$,

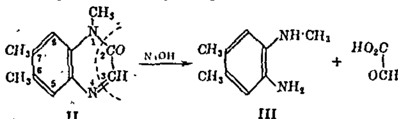




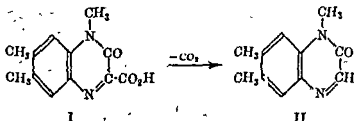
IV III

and since III is $C_9H_{14}N_2$, two carbon and four hydrogen atoms must be accounted for. This can be done by assuming the presence

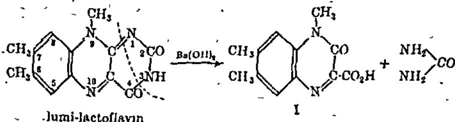
of an ethyl group or of two methyl groups in the benzene ring. Kuhn *et al.* carried out a series of synthetic experiments and showed that III has the structure given, *N*-methyl-4·5-diamino-*o*-xylene. Kuhn then proposed II as the structure of the precursor of III, since this would produce the required products of hydrolysis.



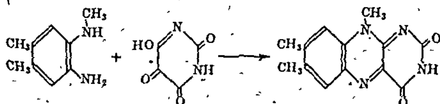
II could therefore have been produced from the β-ketocarboxylic acid I.



Since I and a molecule of urea are obtained from lumi-lactoflavin, the latter could be 6:7:9-trimethyl*iso*alloxazine (6:7:9-trimethylflavin).



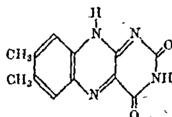
This structure for lumi-lactoflavin has been confirmed by synthesis (Kuhn *et al.*, 1934). *N*-Methyl-4·5-diamino-*o*-xylene is condensed with alloxan hydrate (§2. XVI) in aqueous solution at 50–60°.



Methylation (methyl sulphate) of this synthetic product gives a tetramethyl compound identical with the product obtained by the methylation of the natural lumi-lactoflavin.

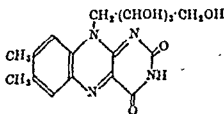
Side-chain of lactoflavin

Exposure of a *neutral* solution of lactoflavin to light produces *lumichrome*, $C_{12}H_{10}O_2N_4$ (Karrer *et al.*, 1934). Analytical work similar to that described for lumi-lactoflavin showed that the structure of lumichrome is



lumichrome

Thus lumichrome is lumi-lactoflavin with a hydrogen atom instead of a methyl group at position 9. This suggests that lactoflavin contains a side-chain (of five carbon atoms) attached to N_9. The Zerewitinoff procedure shows that lactoflavin contains five active hydrogen atoms; thus the molecule contains four hydroxyl groups (one active hydrogen atom is the hydrogen of the NH group at position 3). The presence of these four hydroxyl groups is supported by the fact that the silver salt of lactoflavin (the silver atom replaces the hydrogen of the NH group) forms a tetra-acetate. Thus the side-chain is a tetra-hydroxy derivative, and so a possible structure for lactoflavin is:



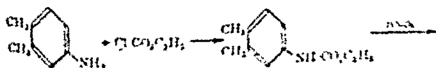
lactoflavin

This side-chain contains three asymmetric carbon atoms, and so there are eight optically active forms possible. Which configuration

is actually present was solved by synthesising a number of pentose derivatives, and it was finally shown by Karrer *et al.* (1935) that the configuration is that of D(−)-ribose. The following syntheses are due to Karrer *et al.* (1935).

(i)

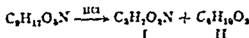
(ii)



Thus lactoflavin is 6:7-dimethyl-9-[D-1′-ribityl]-*iso*alloxazine. Of all the pentoses (and hexoses) used, only the compound from D-ribose shows growth-promoting properties. For this reason vitamin B_2 (lactoflavin) is also known as riboflavin.

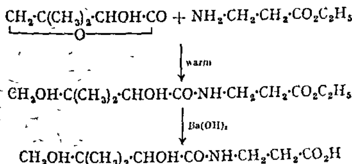
§7. **Pantothenic acid**, $C_9H_{17}O_5N$, is a chick *antidermatitis factor*, and is also capable of promoting the growth of yeast and of bacteria; it has been isolated from many sources, *e.g.*, liver, kidney, yeast, etc.

Pantothenic acid shows the reactions of a monocarboxylic acid, *e.g.*, it can be esterified to form monoesters (R. J. Williams *et al.*, 1939). The application of the method for determining active hydrogen atoms shows that pantothenic acid contains two hydroxyl groups, and since the acid condenses with benzaldehyde (to form a benzylidene derivative) and with acetone (to form an *iso*propylidene derivative), this suggests that the two hydroxy groups are in either the 1:2- or 1:3-position (*cf.* §§8, 9. VII). When warmed with dilute hydrochloric acid, pantothenic acid is hydrolysed into compounds I and II. Investigation of I showed that it was β-alanine



$$C_9H_{17}O_5N \xrightarrow{HCl} \underset{I}{C_3H_7O_2N} + \underset{II}{C_6H_{10}O_3}$$

(actually present as the hydrochloride, $\overset{-}{Cl}\{H_3\overset{+}{N}\cdot CH_2\cdot CH_2\cdot CO_2H\}$).



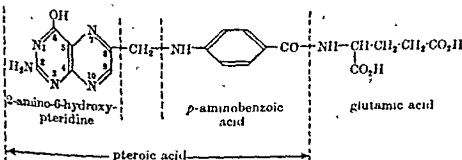
A better yield of pantothenic acid is obtained by warming the dry lactone with the dry sodium salt of β-alanine (R. J. Williams *et al.*, 1940).

§8. **Folic acid complex.** A number of micro-organisms need certain concentrates (prepared from natural sources) for growth, several active principles have been shown to be necessary. In addition to the above property, some of these active principles also exhibit other effects, *e.g.*, the prevention of certain types of anæmia in chicks. The following compounds have been described as constituents of the folic acid complex.

(i) Folic acid. It has been suggested that folic acid from animal sources is different from that from vegetable sources.
(ii) *Lactobacillus casei* factor (three forms).
(iii) *S. lactis* R factor.
(iv) Vitamin B_c factor (this now identified as liver *L. casei* factor).
(v) Vitamin B_c conjugate.
(vi) Vitamins B_{10}, B_{11}, and factors R, S and U.
(vii) Vitamin M.

It is possible that some of these are identical; names have been given by different workers to the active substances that they have isolated (see, *e.g.*, iv).

Angier *et al.* (1946) have shown that liver *L. casei* factor (also called vitamin B_c) is:



fermentation *L. casei* factor contains three glutamic acid residues; yeast vitamin B_c conjugate contains seven glutamic acid residues.

§8a. **Structure of *L. casei* factors** (Angier *et al.*, 1946). The alkaline hydrolysis of the fermentation *L. casei* factor, in the absence of oxygen, formed two molecules of D-glutamic acid and the DL-form of liver *L. casei* factor. On the other hand, the alkaline hydrolysis of the fermentation *L. casei* factor, in the presence of

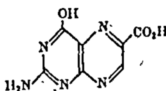



A pteridine B

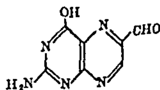
air, gave two substances, I and II. I was shown to be a monocarboxylic acid, and the examination of its ultraviolet absorption spectrum led to the conclusion that it was a pteridine derivative (A ... numbering used here: B is an alternative sy...

A ...

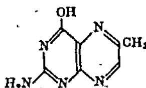
one hydroxyl and one amino group. Oxidation of I with ... water, followed by hydrolysis with hydrochloric acid, produced guanidine, $NH{=}C(NH_2)_2$, as one of the products. The formation of this compound suggests that the amino group is at position 2. Finally, I was shown to be 2-amino-6-hydroxypteridine-8-carboxylic acid by synthesis.



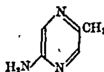
I



IV



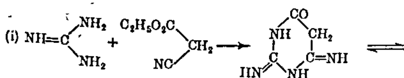
V

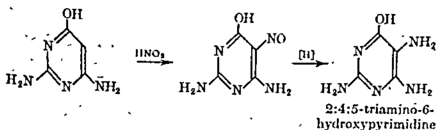


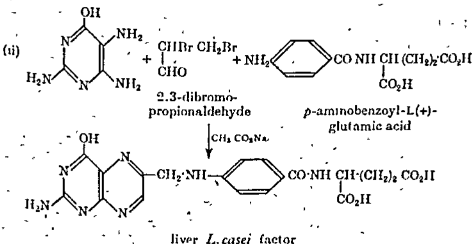
VI

The reactions of compound II showed that it was a primary aromatic amine, and on hydrolysis it gave one molecule of *p*-aminobenzoic acid and three molecules of glutamic acid.

Hydrolysis of the fermentation *L. casei* factor with sulphurous acid gave an aromatic amine, III, and an aldehyde, IV. III, on hydrolysis, gave one molecule of *p*-aminobenzoic acid and three molecules of glutamic acid, *i.e.*, II and III are identical. When the aldehyde IV was allowed to stand in dilute sodium hydroxide solution in the absence of air, compound I and another compound, V, were produced. V, on vigorous hydrolysis, gave 2-amino-5-methylpyrazine, VI. From this it was concluded that V is 2-amino-6-hydroxy-8-methylpteridine, and IV is 2-amino-6-hydroxypteridine-8-aldehyde. Consideration of this evidence led to the suggestion that the liver *L. casei* factor has the structure given in §8; this has been confirmed by synthesis, *e.g.*, that of Angier *et al.* (1946).



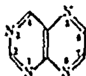




fermentation *L. casei* factor contains three glutamic acid residues; yeast vitamin B_c conjugate contains seven glutamic acid residues.

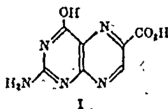
§8a. **Structure of *L. casei* factors** (Angier *et al.*, 1946). The alkaline hydrolysis of the fermentation *L. casei* factor, in the absence of oxygen, formed two molecules of D-glutamic acid and the DL-form of liver *L. casei* factor. On the other hand, the alkaline hydrolysis of the fermentation *L. casei* factor, in the presence of



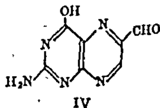


A pteridine B

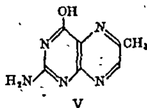
air, gave two substances, I and II. I was shown to be a monocarboxylic acid, and the examination of its ultraviolet absorption spectrum led to the conclusion that it was a pteridine derivative (A is the system of numbering used here; B is an alternative system of numbering frequently used in American publications). A further examination of compound I showed that it also contained one hydroxyl and one amino group. Oxidation of I with chlorine water, followed by hydrolysis with hydrochloric acid, produced guanidine, $NH{=}C(NH_2)_2$, as one of the products. The formation of this compound suggests that the amino group is at position 2. Finally, I was shown to be 2-amino-6-hydroxypteridine-8-carboxylic acid by synthesis.



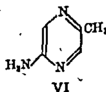
I



IV

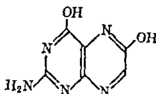


V

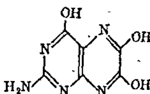


VI

It might be noted, in passing, that the **pterins** are pigments of butterfly wings, wasps, etc.; they were first isolated from butterfly wings.



xanthopterin



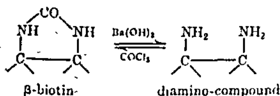
leucopterin

§9. **Biotins** (vitamin H). Bios, an extract of yeast, was shown to be necessary for the growth of yeast (Wildiers, 1901). It was then found that bios consisted of at least two substances (Fulmer *et al.*, 1922), and two years later, Miller showed that three substances were present in bios. The first of these was named Bios I, and was shown to be *meso*inositol (Eastcott, 1928, see also §13). The second constituent, named Bios IIA, was then shown to be β-alanine (Miller, 1936) or pantothenic acid (Rainbow *et al.*, 1939). The third substance, named Bios IIB, was found to be identical with *biotin*, a substance that had been isolated by Kögl *et al.* (1936) as the methyl ester from egg-yolk. Subsequently other factors present in bios have been isolated, *e.g.*, pyridoxin (see §10) and nicotinic acid (§11).

Biotin is a vitamin, being necessary for the growth of animals. In 1940, du Vigneaud *et al.* isolated from liver a substance which had the same biological properties as biotin. Kogl *et al.* (1943) named their extract from egg-yolk α-biotin, and that from liver β-biotin. Both compounds have the same molecular formula $C_{10}H_{16}O_3N_2S$.

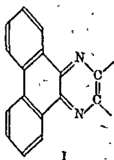
β-Biotin (Bios IIB or biotin), m.p. 230–232°, behaves as a saturated compound (the usual tests showed the absence of an ethylenic double bond). β-Biotin forms a monomethyl ester $C_{11}H_{18}O_3N_2S$ which, on hydrolysis, gives an acid the titration curve of which corresponds to a monocarboxylic acid; thus the formula of β-biotin may be written $C_9H_{15}ON_2S{\cdot}CO_2H$. When heated with barium hydroxide solution at 140°, β-biotin is hydrolysed to carbon dioxide and a diaminocarboxylic acid $C_9H_{18}O_2N_2S$ which, by the action of carbonyl chloride, is reconverted into β-biotin (du Vigneaud *et al.*, 1941). These reactions suggest that β-biotin contains a cyclic ureide structure. Furthermore, since the diaminocarboxylic acid condenses with phenanthraquinone to form a quin-

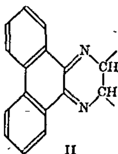
oxaline derivative, it follows that the two amino groups are in the 1:2-positions (*cf.* §19. XII), and thus the cyclic ureide is five-membered. Hence we may write the foregoing reactions as follows:

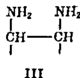


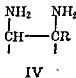
When this diaminocarboxylic acid is oxidised with alkaline permanganate, adipic acid is produced (du Vigneaud *et al.*, 1941). One of the carboxyl groups in adipic acid was shown to be that originally present in *β*-biotin as follows. When the carbomethoxyl group of the methyl ester of *β*-biotin was replaced by an amino group by means of the Curtius reaction (ester → hydrazide → azide → urethan → NH_2; see Vol. I), and the product hydrolysed with barium hydroxide solution, a triamine was obtained which did not give adipic acid on oxidation with alkaline permanganate (du Vigneaud *et al.*, 1941, 1942). It was therefore inferred that *β*-biotin contains a $—(CH_2)_4·CO_2H$ side-chain (*n*-valeric acid side-chain).

The absorption spectrum of the quinoxaline derivative (formed from phenanthraquinone and the diaminocarboxylic acid) showed that it was a quinoxaline, I, and not a dihydroquinoxaline, II; thus the diaminocarboxylic could be III but not IV.









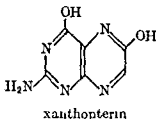
It therefore follows that the *n*-valeric acid side-chain cannot be attached to a carbon atom joined to an amino group.

The nature of the sulphur atom in *β*-biotin was shown to be of the thioether type (*i.e.*, C—S—C) since:

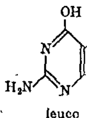
(i) Oxidation of *β*-biotin with hydrogen peroxide produced a sulphone.

(ii) When the methyl ester of *β*-biotin was treated with methyl iodide, a sulphonium iodide was formed.

It might be noted, in passing, that the pter of butterfly wings, wasps, etc.; they were f butterfly wings.



xanthopterin



leuco

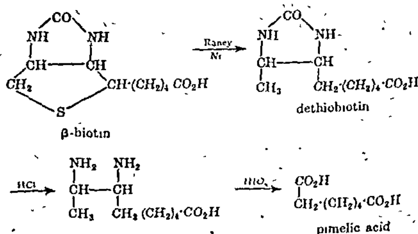
§9. **Biotins** (vitamin H). Bios, an extract to be necessary for the growth of yeast (Wil then found that bios consisted of at least tw *et al.*, 1922), and two years later, Miller sho stances were present in bios. The first of the and was shown to be *meso*inositol (Eastcott, The second constituent, named Bios IIA, wa *β*-alanine (Miller, 1936) or *pantothenic acid* (R The third substance, named Bios IIB, was fo with *biotin*, a substance that had been isolated l as the methyl ester from egg-yolk. Subsequ present in bios have been isolated, *e.g.*, pyrid nicotinic acid (§11).

Biotin is a vitamin, being necessary for the g In 1940, du Vigneaud *et al.* isolated from liver a had the same biological properties as biotin. I named their extract from egg-yolk α-biotin, and *β*-biotin. Both compounds have the same mo $C_{10}H_{16}O_3N_2S$.

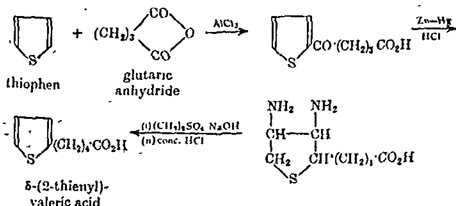
β-Biotin (*Bios IIB or biotin*), m.p. 230–232°, saturated compound (the usual tests showed the ethylenic double bond). *β*-Biotin forms a mon $C_{11}H_{18}O_3N_2S$ which, on hydrolysis, gives an acid curve of which corresponds to a monocarboxylic a formula of *β*-biotin may be written $C_9H_{15}ON_2S$·C heated with barium hydroxide solution at 140°, *β*-bio lysed to carbon dioxide and a diaminocarboxylic acid which, by the action of carbonyl chloride, is reconverted (du Vigneaud *et al.*, 1941). These reactions suggest tl contains a cyclic ureide structure. Furthermore, since tl carboxylic acid condenses with phenanthraquinone to fo

As we have seen, β-biotin does not contain a double bond; hence, from its molecular formula, it was deduced that β-biotin contains two rings (du Vigneaud *et al.*, 1941; Kögl *et al.*, 1941). The sort of argument that may be used is as follows. The molecular formula of β-biotin is $C_{10}H_{16}O_3N_2S$. The carboxyl group may be regarded as a substituent group, and so the parent compound will be $C_9H_{16}ON_2S$. Also, since two NH groups are present, these may be replaced by CH_2 groups; thus the parent compound is $C_{11}H_{18}OS$. The CO group may be replaced by a CH_2 group and the sulphide atom also by a CH_2 group. This gives a compound of formula $C_{12}H_{22}$ which has the same "structure" as β-biotin. Now the formula $C_{12}H_{22}$ corresponds to the general formula C_nH_{2n-2}, and this, for a *saturated* compound, corresponds to a system containing two rings.

When heated with Raney nickel, β-biotin formed *dethiobiotin* by elimination of the sulphur atom (this is an example of the *Mozingo reaction*, 1943). Dethiobiotin, on hydrolysis with hydrochloric acid, gave a diaminocarboxylic acid which, on oxidation with periodic acid, gave pimelic acid (du Vigneaud *et al.*, 1942). These results can be explained by assuming that the sulphur atom is in a five-membered ring and the *n*-valeric acid side-chain is in the position shown.



Further evidence for this structure is given by the fact that the exhaustive methylation of the diaminocarboxylic acid (produced from β-biotin), followed by hydrolysis, gave δ-(2-thienyl)-valeric acid (du Vigneaud *et al.*, 1942); the structure of this compound was confirmed by synthesis.



The above structure for β-biotin has been confirmed by synthesis (Harris *et al.*, 1943, 1944).

$NH_2\cdot CH(CO_2Na)\cdot CH_2SNa$ (Na salt of L-cystine) + $CH_2Cl\cdot CO_2Na$ ⟶ $NH_2\cdot CH(CO_2H)\cdot CH_2\cdot S\cdot CH_2\cdot CO_2H$ $\xrightarrow[(ii)\ CH_3OH-HCl]{(i)\ C_6H_5\cdot COCl}$

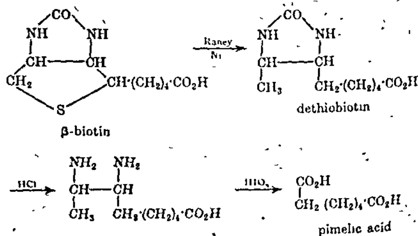
$C_6H_5\cdot CO\cdot NH\cdot CH(CO_2CH_3)\cdot CH_2\cdot S\cdot CH_2\cdot CO_2CH_3$ $\xrightarrow[CH_3OH]{CH_3ONa}$ (ring: $CH(NH\cdot CO\cdot C_6H_5)$—$C(ONa)$=$C(CO_2CH_3)$—S—$CH_2$) $\xrightarrow[CH_3\cdot CO_2H]{HCl}$ (ring: $CH(NH\cdot CO\cdot C_6H_5)$—CO—$CH_2$—S—$CH_2$)

$\xrightarrow[\text{piperidine acetate}]{CHO\cdot(CH_2)_3\cdot CO_2CH_3}$ (ring: $CH(NH\cdot CO\cdot C_6H_5)$—CO—$C$(=$CH\cdot(CH_2)_3\cdot CO_2CH_3$)—S—$CH_2$) $\xrightarrow[(ii)\ Zn-CH_3\cdot CO_2H/(CH_3\cdot CO)_2O]{(i)\ NH_2OH}$

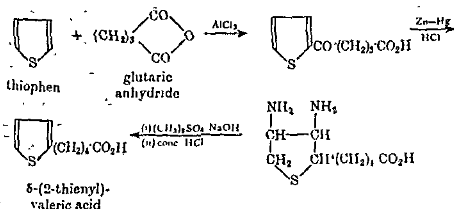
(ring: $CH(NH\cdot CO\cdot C_6H_5)$—$C(NH\cdot CO\cdot CH_3)$=$C\cdot(CH_2)_4\cdot CO_2CH_3$—S—$CH_2$) $\xrightarrow{H_2-Pd}$ (ring: $CH(NH\cdot CO\cdot C_6H_5)$—$CH(NH\cdot CO\cdot CH_3)$—$CH\cdot(CH_2)_4\cdot CO_2CH_3$—S—$CH_2$)

As we have seen, β-biotin does not contain a double bond; hence, from its molecular formula, it was deduced that β-biotin contains two rings (du Vigneaud *et al.*, 1941; Kogl *et al.*, 1941). The sort of argument that may be used is as follows. The molecular formula of β-biotin is $C_{10}H_{16}O_3N_2S$. The carboxyl group may be regarded as a substituent group, and so the parent compound will be $C_9H_{16}ON_2S$. Also, since two NH groups are present, these may be replaced by CH_2 groups; thus the parent compound is $C_{11}H_{18}OS$. The CO group may be replaced by a CH_2 group and the sulphide atom also by a CH_2 group. This gives a compound of formula $C_{12}H_{22}$ which has the same "structure" as β-biotin. Now the formula $C_{12}H_{22}$ corresponds to the general formula C_nH_{2n-2}, and this, *for a saturated* compound, corresponds to a system containing two rings.

When heated with Raney nickel, β-biotin formed *dethiobiotin* by elimination of the sulphur atom (this is an example of the *Mozingo reaction*, 1943). Dethiobiotin, on hydrolysis with hydrochloric acid, gave a diaminocarboxylic acid which, on oxidation with periodic acid, gave pimelic acid (du Vigneaud *et al.*, 1942). These results can be explained by assuming that the sulphur atom is in a five-membered ring and the *n*-valeric acid side-chain is in the position shown.



Further evidence for this struct[illegible] exhaustive methylation of the [illegible] from β-biotin), followed by hy[illegible], gave [illegible] acid (du Vigneaud *et al.*, 1942); the structure of this compound was confirmed by synthesis.



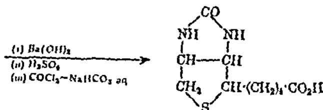
The above structure for β-biotin has been confirmed by synthesis (Harris *et al.*, 1943, 1944).

$NH_2 \cdot CH(CO_2Na) \cdot CH_2SNa$ (Na salt of L-cystine) + $CH_2Cl \cdot CO_2Na$ ⟶ $NH_2 \cdot CH(CO_2H) \cdot CH_2 \cdot S \cdot CH_2 \cdot CO_2H$ $\xrightarrow[(ii)\ CH_3OH-HCl]{(i)\ C_6H_5 \cdot COCl}$

$C_6H_5 \cdot CO \cdot NH \cdot CH(CO_2CH_3) \cdot CH_2 \cdot S \cdot CH_2 \cdot CO_2CH_3$ $\xrightarrow[CH_3OH]{CH_3ONa}$ (ring: $NH \cdot CO \cdot C_6H_5$ — CH—CONa=C(CO_2CH_3), CH_2, S) $\xrightarrow[CH_3 \cdot CO_2H]{HCl}$ (ring: $NH \cdot CO \cdot C_6H_5$ — CH—CO, CH_2, CH_2, S)

$\xrightarrow[\text{piperidine acetate}]{CHO \cdot (CH_2)_3 \cdot CO_2CH_3}$ (ring: $NH \cdot CO \cdot C_6H_5$ — CH—CO, CH_2, $C{=}CH \cdot (CH_2)_3 \cdot CO_2CH_3$, S) $\xrightarrow[(ii)\ Zn-CH_3 \cdot CO_2H/(CH_3 \cdot CO)_2O]{(i)\ NH_2OH}$

(ring: $C_6H_5 \cdot CO$—NH—CH, $CO \cdot CH_3$—NH—C, CH_2, $C \cdot (CH_2)_4 \cdot CO_2CH_3$, S) $\xrightarrow{H_2-Pd}$ (ring: $C_6H_5 \cdot CO$—NH—CH, $CO \cdot CH_3$—NH—CH, CH_2, $CH \cdot (CH_2)_4 \cdot CO_2CH_3$, S)



Two racemates were isolated, one of which was ($\pm$)-β-biotin; this was resolved *via* its esters with ($-$)-mandelic acid.

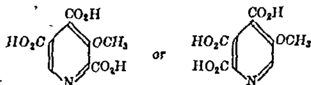
Examination of the β-biotin formula shows the presence of three asymmetric carbon atoms; the rings are fused in the *cis*-position in β-biotin.

The structure of α-biotin is uncertain.

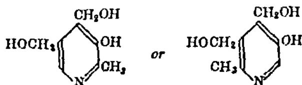
§10. **Pyridoxin** (Adermin, vitamin B_6), $C_8H_{11}O_3N$, is obtained from rice bran and yeast; it cures dermatitis in rats. Pyridoxin behaves as a weak base, and the usual tests showed the absence of methoxyl and methylamino groups. Application of the Zerewitinoff method showed the presence of three active hydrogen atoms. When treated with diazomethane, pyridoxin formed a monomethyl ether which, on acetylation, gave a diacetyl derivative (Kuhn *et al.*, 1938). It therefore appears that the three oxygen atoms in pyridoxin are present as hydroxyl groups, and since one is readily methylated, this one is probably phenolic. This conclusion is supported by the fact that pyridoxin gives the ferric chloride colour reaction of phenols. Thus the other two hydroxyl groups are alcoholic.

Examination of the ultraviolet absorption spectrum of pyridoxin showed that it is similar to [illegible] s
therefore inferred that pyrido[illegible] s
phenolic group in position 3. [illegible] l
on the monomethyl ether of pyridoxin, this leads to the conclusion that the two alcoholic groups are not on adjacent carbon atoms in a side-chain (Kuhn *et al.*, 1939). When this methyl ether is *very carefully* oxidised with alkaline potassium permanganate, the product is a methoxypyridinetricarboxylic acid, $C_9H_7O_7N$. This acid gave a blood-red colour with ferrous sulphate, a reaction which is characteristic of pyridine-2-carboxylic acid; thus one of the three carboxyl groups is in the 2-position. When the methyl ether of pyridoxin was oxidised with alkaline permanganate under the usual conditions, the products were carbon dioxide and the anhydride of a dicarboxylic acid, $C_8H_5O_4N$; thus these two carboxyl groups are in the *ortho*-position. Furthermore, since this anhydride,

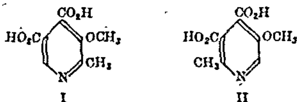
on hydrolysis to its corresponding acid, did not give a red colour with ferrous sulphate, there is no carboxyl group in the 2-position. It therefore follows that, on decarboxylation, the tricarboxylic acid eliminates the 2-carboxyl group to form the anhydride; thus the tricarboxylic acid could have either of the following structures.



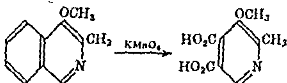
Now pyridoxin methyl ether contains three oxygen atoms (one as methoxyl and the other two alcoholic); it is therefore possible that two carboxyl groups in the tricarboxylic acid could arise from two CH_2OH groups, and the third from a methyl group, *i.e.*, pyridoxin could be either of the following:



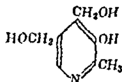
A decision between the two structures was made on the following evidence. When pyridoxin methyl ether was oxidised with barium permanganate, the product was a dicarboxylic acid, $C_9H_9O_5N$, which did not give a red colour with ferrous sulphate; thus there is no carboxyl group in the 2-position. Also, since the dicarboxylic acid formed an anhydride and gave a phthalein on fusion with resorcinol, the two carboxyl groups must be in the *ortho*-position. Furthermore, analysis of both the dicarboxylic acid and its anhydride showed the presence of a methyl group. Thus the structure of this dicarboxylic acid is either I or II.



Kuhn *et al.* (1939) showed that the anhydride was that of I from its formation by the oxidation of 4-methoxy-3-methyl-*iso*quinoline (a synthetic compound of known structure).



Hence, on the foregoing evidence, pyridoxin is



pyridoxin

This structure has been confirmed by synthesis, *e.g.*, that of Harris and Folkers (1939):

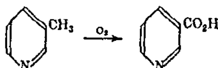
ethoxyacetyl-acetone cyano-acetamide

pyridoxin

§11. Nicotinic acid and nicotinamide. These two compounds have been shown to be the human pellagra-preventing (P.P.) factor. Nicotinamide is part of the co-enzymes codehydrogenase I and II, which play a part in many biological oxidations.

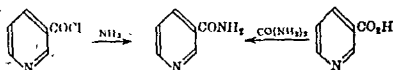
Nicotinic acid (*Niacin*) was first prepared by the oxidation of nicotine (§21. XIV). This is now used as a commercial method;

another commercial method for the preparation of nicotinic acid is the vapour-phase oxidation of 3-methylpyridine (β-picoline) in the presence of a vanadium and iron catalyst.

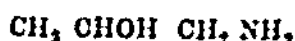


Still another commercial method is the oxidation of quinoline to quinolinic acid, which is then decarboxylated to nicotinic acid (see also §21. XIV).

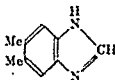
Nicotinamide, m.p. 131°, is manufactured by various methods, *e.g.*, by the action of ammonia on nicotinyl chloride, or by heating nicotinic acid with urea in the presence of a molybdenum catalyst.



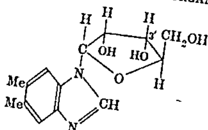
§12. **Vitamin B_{12}, Cyanocobalamin.** This is the anti-pernicious anæmia factor, and has been isolated from liver extract. Folic acid (§8) also has anti-anæmic properties. Vitamin B_{12} has been obtained as a red crystalline substance (Folkers *et al.*, 1948; Smith *et al.*, 1948, 1949), and the elements present have been shown to be C, H, O, N, P, Co; this vitamin is the first natural product found to contain cobalt. The cobalt has been shown to be attached to a cyano group. The hydrolysis of vitamin B_{12} with hydrochloric acid under different conditions produces ammonia, 1-aminopropan-2-ol (I), 5.6-dimethylbenziminazole (II), 5.6-dimethylbenziminazole 1-α-D-ribofuranoside (III), and the 3'-phosphate of III (Folkers *et al.*, 1949, 1950; Todd *et al.*, 1950). Compound IV (a succinimide derivative) has also been isolated by the chromic acid oxidation of hydrolysed vitamin B_{12} (Folkers, 1955).

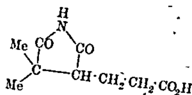


I



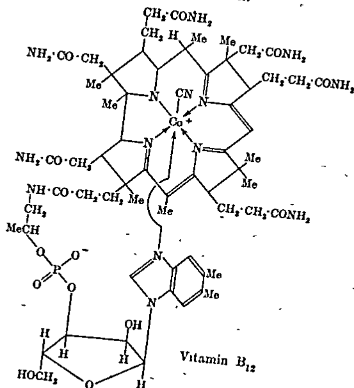
II





III IV

Other work has shown that six amido groups are present in the molecule. Also, alkaline hydrolysis of vitamin B_{12} gives a mixture consisting mainly of a penta- and a hexacarboxylic acid, in both of which the nucleotide fragment is absent. As the result of a detailed X-ray analysis of the hexacarboxylic acid, vitamin B_{12} has been assigned the structure shown.



Vitamin B_{12}

A point of interest is that the arrangement of the four pyrrole nuclei is somewhat similar to that in the natural porphin derivatives such as hæm and chlorophyll (§§2, 7. XIX).

A number of vitamin B_{12} compounds have now been isolated which differ only in the nature of the basic component of the nucleotide. The remainder of the molecule, which is referred to as Factor B, is common to all the members of the vitamin B_{12} group.

§13. Other compounds of the vitamin B complex. Three other compounds which have definitely been isolated from the vitamin B complex are:

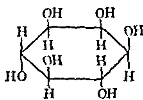
(i) *p*-Aminobenzoic acid; this is a growth factor for bacteria.

(ii) *meso*Inositol (m.p. 225–226°). This is a growth factor in animals, and its configuration has been elucidated by Posternak (1942; see also §11 iv. IV).

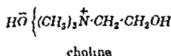
(iii) Choline. The absence of this compound leads to the formation of a fatty liver in animals.



p-amino-benzoic acid



*meso*inositol



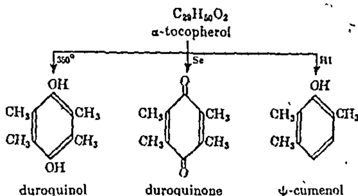
choline

Other vitamins of the vitamin B complex that have been said to exist are vitamins B_3, B_4, B_5, B_{10}, B_{11}, B_{13}, B_{14}, and others.

VITAMIN E GROUP

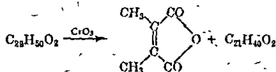
§14. Introduction. *Vitamin E is the anti-sterility factor; it* occurs in seed germ oils. It is now known that there are three closely related compounds comprising "vitamin E"; all three are biologically active, and are known as *α-, β- and γ-tocopherol.* The main source of α- and β-tocopherol is wheat germ oil; the γ-compound is obtained from cotton seed oil. Wheat germ oil was first subjected to chromatographic analysis to remove sterols, etc., and then the α- and β-tocopherols were purified by conversion into their crystalline allophanates (see §12, XII) or 3:5-dinitrobenzoates. Hydrolysis of these derivatives gave the tocopherols as pale yellow oils.

§15. **α-Tocopherol, $C_{29}H_{50}O_2$.** When α-tocopherol is heated at 350°, duroquinol is obtained (Fernholz, 1937). On the other hand, when heated with selenium, α-tocopherol forms duroquinone (McArthur *et al.*, 1937). Finally, when heated with hydriodic acid, ψ-cumenol is formed (John *et al.*, 1937).

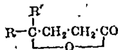


The formation of these products led to the suggestion that α-tocopherol was the monoether of duroquinol; the possibility that it might be the diether was ruled out by the fact that α-tocopherol forms an allophanate, which indicates the presence of one free hydroxyl group. This monoether structure was shown to be incorrect by the fact that the ultraviolet absorption spectra of various monoethers of duroquinol were different from that of α-tocopherol (Fernholz, 1938).

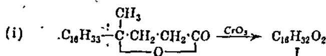
Oxidation of α-tocopherol with chromic acid forms dimethylmaleic anhydride and a compound $C_{21}H_{40}O_2$.

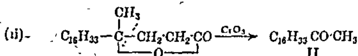


This latter compound was shown to be an optically active saturated lactone. This lactone was then shown to be derived from a γ-hydroxyacid in which the hydroxyl group is tertiary, *e.g.*, the acid lactonised immediately its salt was acidified, and also could not be oxidised to a keto-acid. Thus the structure of this lactone may be written (R + R′ = 17C):

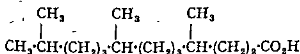


Now α-tocopherol acetate, on oxidation with chromic acid, forms an acid, $C_{16}H_{32}O_2$, I, and a ketone, $C_{18}H_{36}O$, II. Both of these compounds must be produced by the oxidation of the lactone at different points in the chain. Fernholz therefore suggested that if in the lactone $R = C_{16}H_{33}$ and $R' = CH_3$, then the products I and II can be accounted for; thus:

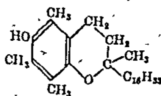




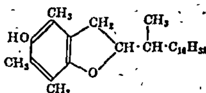
Fernholz then showed that the acid (I) contained methyl groups (*cf.* §3. IX), and was led to propose a structure based on the isoprene unit, *viz.*,



The evidence obtained so far indicates the presence of a substituted benzene ring and a long side-chain in α-tocopherol. When the monoethers of duroquinol (see above) were oxidised with silver nitrate solution, the action took place far more slowly than for α-tocopherol when oxidised under the same conditions. Furthermore, whereas the former compounds were oxidised to duroquinone, the latter compound gave a red oil which appeared to have approximately the same molecular weight as α-tocopherol (Fernholz, 1938). Since duroquinone is not split off during this oxidation, it suggests that the side-chain is connected to the aromatic ring by a carbon bond as well as an ether link. In this case α-tocopherol is either a chroman or coumaran derivative:



chroman structure



coumaran structure

According to Fernholz, the oxidation products are best explained on the chroman structure. This has been supported by ultraviolet absorption measurements of α-tocopherol (John *et al.*, 1938).

Karrer *et al.* (1938) have synthesised (±)-α-tocopherol by condensing trimethylquinol with phytyl bromide (§30. VIII).

$$\text{trimethylquinol} + BrCH_2{-}CH{=}C(CH_3){-}C_{16}H_{33} \xrightarrow[\text{light petrol}]{ZnCl_2} [\text{intermediate}] \longrightarrow$$

(±)-α-tocopherol

This synthesis, however, is not completely unambiguous, since phenols may condense with allyl compounds to form coumarans. Smith *et al.* (1939) have shown that γ:γ-disubstituted halides form only chromans, and since phytyl bromide is a halide of this type, this strengthens the course of the synthesis given above. Finally, Smith *et al.* (1942) have carried out an unambiguous synthesis of α-tocopherol as follows:

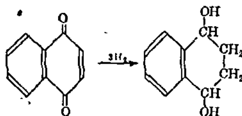
(i) $\xrightarrow[\text{(ii) Mg}]{\text{(i) } PBr_3}$ $CH_2{\cdot}CH_2MgBr$

(ii) $CH_2{\cdot}CH_2MgBr$ + $CO{\cdot}(CH_2)_3{\cdot}CH(CH_3){\cdot}(CH_2)_3{\cdot}CH(CH_3){\cdot}(CH_2)_3{\cdot}CH(CH_3)_2$ (with CH_3 on the CO carbon)

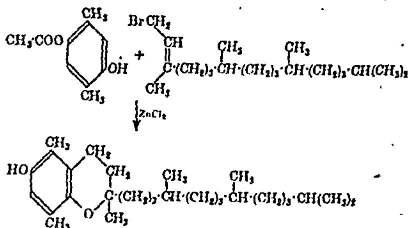
VITAMIN K GROUP

§19. Introduction. Dam *et al.* (1939) and Doisy *et al.* (1939) isolated vitamin K from alfalfa, and called it vitamin K_1 to distinguish it from a substance called vitamin K_2 which had been isolated from putrefied fish meal by Doisy *et al.* (1939). Both are *antihæmorrhagic vitamins; they are connected with the enzymes* involved in blood clotting, a deficiency of them lengthening the time of blood clotting.

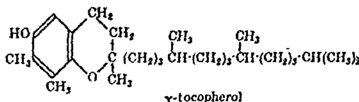
§20. Vitamin K_1 (α-phylloquinone), $C_{31}H_{46}O_2$, is a light yellow oil. The redox potential of vitamin K_1 is very similar to that of 1:4-quinones (Karrer *et al.*, 1939), and its absorption spectrum is very similar to that of 2:3-disubstituted 1:4-naphthaquinones (McKee *et al.*, 1939). Thus vitamin K_1 appears to be a 1:4-naphthaquinone derivative, and this is in keeping with the fact that the vitamin is very sensitive to light and to alkalis. Now the catalytic hydrogenation of vitamin K_1 causes the addition of four molecules of hydrogen (McKee *et al.*, 1939); the product is a colourless compound. *Since it is known that three molecules of hydrogen are* added when 1:4-naphthaquinone is reduced under these conditions, the addition of a fourth molecule of hydrogen to the vitamin suggests the presence of an ethylenic double bond in a side-chain.



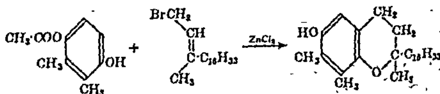
When subjected to reductive acetylation (*i.e.*, acetylated under reducing conditions), vitamin K_1 is converted into the diacetate of dihydrovitamin K_1 (Binkley *et al.*, 1939). This diacetate is difficult to hydrolyse; this is a property characteristic of 2:3-disubstituted 1:4-naphthaquinones. When oxidised with chromic acid, vitamin K_1 gives phthalic acid, but when the oxidation is carried out under controlled conditions, the product is a compound with the molecular formula $C_{13}H_{10}O_4$. This latter compound was subsequently shown to be 2-methyl-1:4-naphthaquinone-3-acetic acid (Binkley *et al.*, 1939).



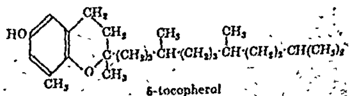
§17. γ-**Tocopherol**, $C_{28}H_{48}O_2$. This is isomeric with β-tocopherol; the only difference is the positions of the two methyl groups in the benzene ring, *e.g.*, when heated with hydriodic acid, γ-tocopherol gives *o*-xyloquinol. Thus γ-tocopherol is


γ-tocopherol

This structure has been confirmed by synthesis, starting from the monoacetate of *o*-xyloquinol and phytyl bromide.



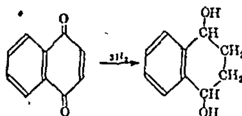
§18. δ-**Tocopherol**, $C_{27}H_{46}O_2$. This was isolated from soyabean oil by Stern *et al.* (1947); it is a yellow oil, and is inactive physiologically. The structure of δ-tocopherol is


δ-tocopherol

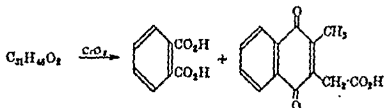
VITAMIN K GROUP

§19. Introduction. Dam *et al.* (1939) and Doisy *et al.* (1939) isolated vitamin K from alfalfa, and called it vitamin K_1 to distinguish it from a substance called vitamin K_2 which had been isolated from putrefied fish meal by Doisy *et al.* (1939). Both are antihæmorrhagic vitamins, they are connected with the enzymes involved in blood clotting, a deficiency of them lengthening the time of blood clotting.

§20. Vitamin K_1 (α-phylloquinone), $C_{31}H_{46}O_2$, is a light yellow oil. The redox potential of vitamin K_1 is very similar to that of 1:4-quinones (Karrer *et al.*, 1939), and its absorption spectrum is very similar to that of 2·3-disubstituted 1.4-naphthaquinones (McKee *et al.*, 1939). Thus vitamin K_1 appears to be a 1 4-naphthaquinone derivative, and this is in keeping with the fact that the vitamin is very sensitive to light and to alkalis. Now the catalytic hydrogenation of vitamin K_1 causes the addition of four molecules of hydrogen (McKee *et al.*, 1939); the product is a colourless compound. Since it is known that three molecules of hydrogen are added when 1.4-naphthaquinone is reduced under these conditions, the addition of a fourth molecule of hydrogen to the vitamin suggests the presence of an ethylenic double bond in a side-chain.

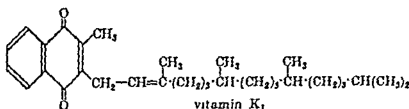


When subjected to reductive acetylation (*i.e.*, acetylated under reducing conditions), vitamin K_1 is converted into the diacetate of dihydrovitamin K_1 (Binkley *et al.*, 1939). This diacetate is difficult to hydrolyse; this is a property characteristic of 2·3-disubstituted 1.4-naphthaquinones. When oxidised with chromic acid, vitamin K_1 gives phthalic acid, but when the oxidation is carried out under controlled conditions, the product is a compound with the molecular formula $C_{13}H_{10}O_4$. This latter compound was subsequently shown to be 2-methyl-1:4-naphthaquinone-3-acetic acid (Binkley *et al.*, 1939).

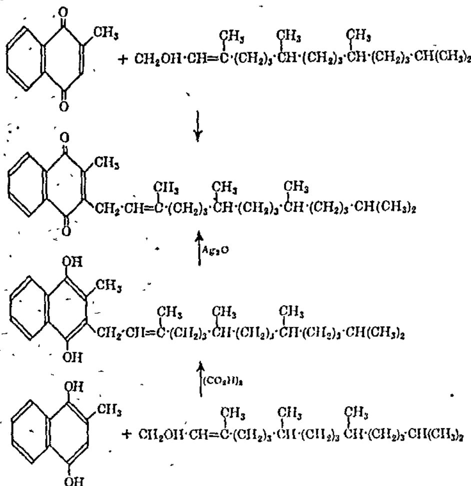


Thus the presence of the 1:4-naphthaquinone structure is confirmed, and at the same time these products show that one ring is unsubstituted and that the other (the quinonoid ring) has substituents in the 2- and 3-positions.

When the diacetate of dihydrovitamin K_1 (see above) was subjected to ozonolysis, a compound $C_{18}H_{36}O$ was obtained, which was then shown to be identical with the ketone produced by the oxidation of phytol (McKee *et al.*, 1939; *cf.* Smith's synthesis of α-tocopherol, §15). Hence, on the evidence obtained above, vitamin K_1 is 2-methyl-3-phytyl-1:4-naphthaquinone.



This structure has been confirmed by synthesis: Almquist *et al.* (1939) obtained vitamin K_1 by condensing 2-methyl-1:4-naphthaquinone with phytol; Fieser *et al.* (1939) obtained a better yield by heating 2-methyl-1,4-naphthaquinol with phytol in dioxan solution in the presence of anhydrous oxalic acid, and then oxidising the product, dihydrovitamin K_1, with silver oxide in ether.

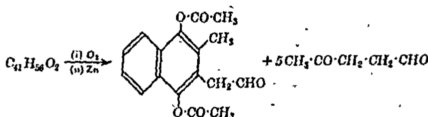


Wendler *et al.* (1954) have obtained vitamin K_1 in good yield by condensing the 1-acetyl derivative of 2-methyl-1:4-naphthaquinol with phytol in the presence of boron trifluoride.

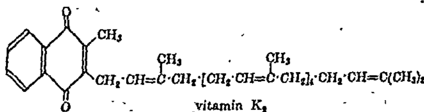
§21. Vitamin K_2, $C_{41}H_{56}O_2$, is a yellow solid, m.p. 54°; it is less potent than vitamin K_1. It was shown to contain a 1:4-naphthaquinone nucleus by the facts that it is sensitive to light and to alkalis, and that it has an absorption spectrum similar to that of vitamin K_1 (McKee *et al.*, 1939). When catalytically reduced, vitamin K_2 adds on nine molecules of hydrogen, and since three of these are absorbed by the naphthaquinone nucleus (see §20), it therefore suggests that there is a side-chain present which contains six double bonds. Furthermore, since vitamin K_2 does not form an adduct with maleic anhydride, no conjugation is present (McKee *et al.*, 1939).

That these six double bonds are ethylenic is shown by the fac that on reductive acetylation, vitamin K_2 forms the diacetate o dihydrovitamin K_2, which can add on six molecules of bromine

The oxidation of vitamin K_2 with permanganate produces phthalic acid; therefore one ring is unsubstituted. On the other hand, when ozone is passed into a solution of vitamin K_2 in acetic acid, and the product then treated with zinc dust in ether, 1:4-diacetoxy-2-methylnaphthalene-3-acetaldehyde is produced. At the same time there is obtained lævulaldehyde in a yield of 93 per cent. calculated on the basis that one molecule of vitamin K_2 can produce five molecules of the aldehyde.



Acetone is also formed in this reaction, and is obtained in a yield of 56 per cent. based on the assumption that one molecule of acetone is produced from one molecule of vitamin K_2 (McKee *et al.*, 1940). On this evidence, it has been suggested that vitamin K_2 is 3-difarnesyl-2-methyl-1:4-naphthaquinone (Binkley *et al.*, 1940).



§22. **Other compounds possessing antihæmorrhagic properties.** It has been shown that simple 1:4-naphthaquinones have blood-clotting properties. 2-Methyl-1:4-naphthaquinone is more active than either vitamin K_1 or K_2 (Fernholz *et al.*, 1939); it is therefore used instead of the natural vitamins. *Phthiocol* (3-hydroxy-2-methyl-1:4-naphthaquinone) is also an active compound, and is water-soluble. It is also interesting to note that many quinones other than 1:4-naphthaquinones have also been found to be active, *e.g.*, some *p*-benzoquinones.

Reading References

Vitamins, A Survey of Recent Knowledge, Medical Research Council Report (1932).
Rosenberg, *Chemistry and Physiology of the Vitamins*, Interscience Publishers (1942).
Vitamins and Hormones, Academic Press (1943–).
Stewart and Graham, *Recent Advances in Organic Chemistry*, Longmans, Green. Vol. III (1948, 7th ed.). Ch 2 Vitamins.
Robinson, *The Vitamin B Complex*, Chapman and Hall (1951)
Todd, Vitamins of the B Group, *J.C.S.*, **1941**, 427.
Smith, The Chemistry of Vitamin E, *Chem. Reviews*, 1940, **27**, 287.
Chu, Biotin and Related Compounds, *Chem. and Ind.*, **1948**, 115.
Ann. Reports (Chem. Soc.). β-Biotin; 1943, **40**, 172; 1944, **41**, 215; 1946, **43**, 239; 1948, **45**, 220.
Ann. Reports (Chem. Soc). Pterins; 1946, **43**, 250; 1948, **45**, 226; 1950, **47**, 241; 1951, **48**, 226.
Albert, The Pteridines, *Quart. Reviews (Chem Soc)*, 1952, **6**, 197
Editorial Report on Nomenclature, *J.C.S.*, **1951**, 3526. Nomenclature of the Vitamins.
Harris, *Vitamins in Theory and Practice*, Cambridge University Press (1955, 4th ed.).
Structure of Vitamin B_{12}.
(i) Dorothy Crowfoot Hodgkin *et al.*, *Nature*, 1955, **176**, 325; 1956, **178**, 64.
(ii) Todd *et al.*, *Nature*, 1955, **176**, 330.
(iii) *Chem. Soc. Special Publ.* No. 3, 1955.

CHAPTER XVIII

CHEMOTHERAPY

§1. **Introduction.** The term *chemotherapy* was introduced by Ehrlich (1909), and it now appears to be used in the sense of the treatment of diseases due to bacterial invasion by chemical compounds which destroy the micro-organisms without affecting, to any material extent, the tissues (of the host). Many compounds, *e.g.*, formaldehyde, phenol, iodine, etc., are also active in destroying bacteria. These compounds, however, are applied *externally*, and tend to destroy the tissues; thus they are not included under the heading of therapeutic agents, but are known as *disinfectants*.

The first compounds to be used by Ehrlich (1907) were organic dyes. From then onwards, organic compounds of diverse chemical structures have been used in chemotherapy. It has now been found that a given compound is specific in its toxicity towards a particular micro-organism. The relationship between chemical structure and chemotherapeutic action is extremely complicated, but some progress has been made in this field.

Compounds which exert various physiological effects of therapeutic value are collectively known as *drugs*. The ideal requirement of a drug is that, on administration (to the host), it should be localised at the site where it is required. In practice, however, no drug behaves in this way, but tends to distribute itself anywhere in the tissues of the host. Another difficulty is that cells, which were originally susceptible to a particular drug, may acquire a tolerance (resistance) to that drug. In some cases it has been found that the drug actually reverses its original action, *i.e.*, it stimulates the cell instead of inhibiting it.

There have been three approaches to the problem of finding a drug to combat a particular disease:—

(i) The method of trial and error. This involves the trial of all kinds of compounds, natural and synthetic.

(ii) The method requiring a knowledge of the cell system, and then synthesising compounds which interfere with it.

(iii) The method in which one starts with a compound known to have some of the required activity (this information has been gained from the previous methods), and then to vary the structure of the molecule systematically. This method has, so far, proved to be the most fruitful.

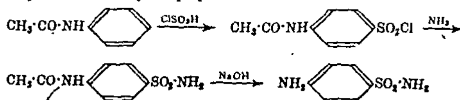
§2. Sulphonamides. Sulphanilamide (*p*-aminobenzenesulphonamide) and its derivatives have great antibacterial powers; sulphanilamide itself is widely used in medicine against "cocci infections"—streptococci, gonococci and pneumococci. Research in the sulphonamide field was stimulated by the discovery of Domagk (1934) that prontosil (see below) had a curative effect ... cci.

... ents of the amide ..., and substituents of the amino group are called N^4-substituents.

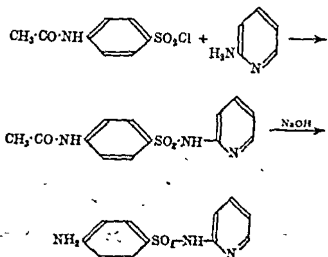


sulphanilamide

Sulphanilamide may be prepared from acetanilide:

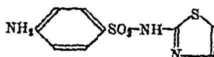


Sulphapyridine (N^1-2-pyridylsulphanilamide) was the first drug to effect cures of pneumonia; it is more potent than sulphanilamide. It may be prepared as follows:



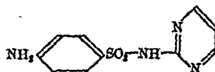
This compound was introduced under the trade name of M

Sulphathiazole (N^1-2-thiazolylsulphanilamide) is more potent than *Sulphapyridine* and less toxic; it is used mainly in severe infections. It is prepared in the same way as *Sulphapyridine*



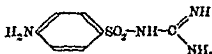
except that 2-aminothiazole is used instead of 2-aminopyridine.

Sulphadiazine (N^1-2-pyrimidylsulphanilamide; *Sulphapyrimidine*) is less toxic than *Sulphathiazole*; it is the most widely used of the "sulpha" drugs, its main use being for mild infections. It is prepared in the same way as the previous compound, except that 2-aminopyrimidine is used in this case.

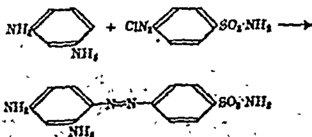


Sulphamezathine(N'-2(4:6-dimethylpyrimidyl)-sulphanilamide) is also used for general purposes.

Sulphaguanidine, since it is only slightly absorbed in the intestinal tract, can therefore be given in relatively large doses in the treatment of bacillary dysentery.

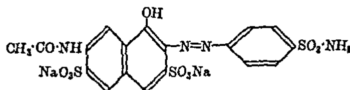


Prontosil (4-sulphonamido-2':4'-diaminoazobenzene) was the first sulphonamide to be used in medicine. It is prepared by diazotising sulphanilamide and then coupling with *m*-phenylenediamine.

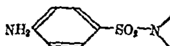


It was suggested that *Prontosil* broke down in the body to sulphanilamide; this led to the discovery that the latter compound is very active against bacteria.

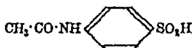
Prontosil S is more soluble than *Prontosil*.



Mechanism of action of the sulphonamides. It appears that the antibacterial activity of the sulphonamides is associated with the group



Some compounds containing slight variations from this structure are also active, *e.g.*,

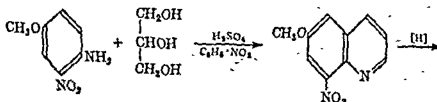


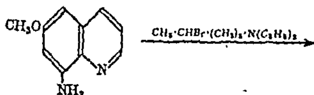
Compounds in which the amino group is *ortho* or *meta* to the sulphonamido group are either less active or completely inactive.

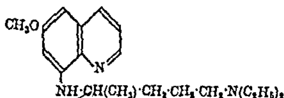
p-Aminobenzoic acid is an essential growth factor for most bacteria susceptible to the sulphonamides. The theory of action is that, owing to the similarity in structure, bacteria absorb a sulphonamide "by mistake", and once this compound is ingested, the bacteria cease to grow in numbers (Woods, 1940). Thus the sulphonamides are not bactericidal but bacteriostatic.

§3. Antimalarials. Quinine (§25b. XIV) was originally the only drug known to be effective against malaria. Now there is a number of synthetic compounds used for this purpose, *e.g.*, *Plasmoquin*, *Mepacrine*, *Proguanil*.

Plasmoquin (*Pamaquin*) is 8-(4'-diethylamino-1'-methylbutylamino)-6-methoxyquinoline. One preparation that has been described for this compound is the condensation between 4-bromo-1-diethylaminopentane and 8-amino-6-methoxyquinoline, the latter being prepared from 4-amino-3-nitroanisole by means of the Skraup synthesis (see Vol. I).

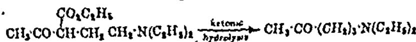




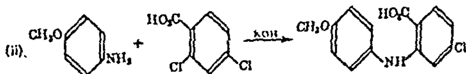


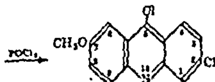
Mepacrine (*Atebrin, Quinacrine*) is 2-chloro-5-(4'-diethylamino-1'-methylbutylamino)-7-methoxyacridine. It is better than quinine, and it has been prepared as follows:

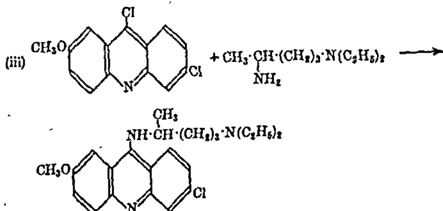




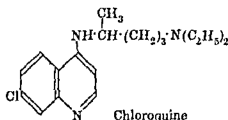




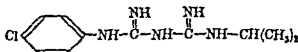




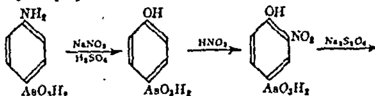
Mepacrine has certain unpleasant side-effects (such as producing a yellow colour in the skin, nausea, etc.), and a drug superior to both quinine and *Mepacrine* is *Chloroquine* (*Aralen*).

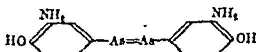


Proguanil (*Paludrine*) is N^1-*p*-chlorophenyl-N^5-*iso*propyl-diguanide. It is superior to *Mepacrine* and *Chloroquine*, and appears to be the best anti-malarial known at the present time.

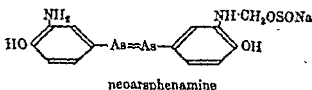


§4. **Arsenical drugs.** A particularly important use of arsenical drugs is in the trea [illegible]
Arsphenamine [illegible]
Ehrlich (1909); it [illegible]
and may be prepared as follows:





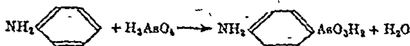
Arsphenamine is an unstable compound; it is stable as its dihydrochloride which, however, cannot be used as such but must be converted into the soluble sodium salt. Ehrlich (1912) overcame this difficulty by preparing **neoarsphenamine** (*Neosalvarsan*), a



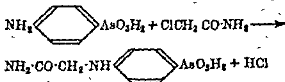
neoarsphenamine

soluble compound, which may be produced by condensing arsphenamine with sodium formaldehydesulphoxylate, $CH_2OH{\cdot}SO_2Na$.

Atoxyl is the sodium salt of *p*-arsanilic acid (*p*-aminophenylarsonic acid); it is used in the treatment of sleeping sickness. *p*-Arsanilic acid may be prepared by heating aniline with arsenic acid at 200° (*cf.* sulphanilic acid, Vol. I).



Tryparsamide is the sodium salt of *N*-phenylglycineamide-*p*-arsonic acid; it is less toxic than *Atoxyl*, and may be prepared by refluxing the latter with chloroacetamide.



§5. **Antibiotics.** Many micro-organisms produce within themselves chemical substances which, when excreted, interfere with the growth or metabolism of other micro-organisms. Such compounds are known as *antibiotics*, and need be present only in low concentration to bring about this antibiotic action. Antibiotics are thus chemotherapeutic agents.

In 1929, Fleming discovered a mould of the *Penicillium* species which inhibited the growth of certain bacteria. This observation was investigated later by a number of workers and culminated in the isolation of the active principle *penicillin*. At the same time, research along this line led to the isolation of many other antibiotics.

§6. The penicillins. Penicillin is the name given to the mixture of natural compounds having the molecular formula $C_9H_{11}O_4N_2SR$, and differing only in the nature of R. There are at least five natural penicillins.

Chemical Name	Other Names	R
Pent-2-enylpenicillin	Penicillin-I or F	$—CH_2·CH{=}CH·CH_2·CH_3$
Benzylpenicillin	Penicillin-II or G	$—CH_2·C_6H_5$
p-Hydroxybenzylpenicillin	Penicillin-III or X	$—CH_2·C_6H_4·OH(1\ 4)$
n-Heptylpenicillin	Penicillin-K	$—(CH_2)_6·CH_3$
n-Amylpenicillin	Dihydro-F-penicillin	$—(CH_2)_4·CH_3$

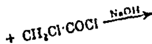
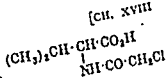
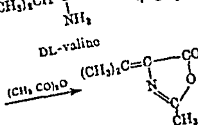
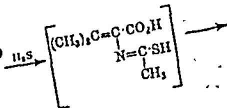
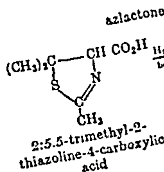
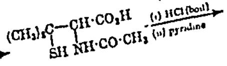
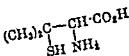
Commercial preparations of penicillin contain one or more of the penicillins in varying proportions. *It has been found that the* addition to the culture medium of various compounds containing a benzyl group, *e.g.*, phenylacetic acid, phenylacetamide, etc., increases the total yield of penicillin, and also the proportion of benzylpenicillin. Similarly, the addition of compounds containing the *p*-hydroxybenzyl group to the culture medium increases the proportion of *p*-hydroxybenzylpenicillin. On the other hand, by adding various compounds to the culture medium, a number of "unnatural" penicillins have been prepared.

§6a. Structure of the penicillins. The penicillins are all strong monobasic acids, *e.g.*, they form salts. The penicillins are hydrolysed by hot dilute inorganic acids; one carbon atom is eliminated as carbon dioxide, and two products are obtained in equimolecular amounts, one being an amine, *penicillamine*, and the other an aldehyde, *penilloaldehyde*. All the penicillins give the same amine, but different aldehydes; it is the latter which contain the R group.

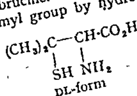
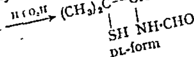
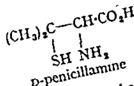


$$C_9H_{11}O_4N_2SR + 2H_2O \xrightarrow{HCl} CO_2 + C_5H_{11}O_2NS + C_3H_4O_2NR$$

D-Penicillamine, $C_5H_{11}O_2NS$. *Since penicillamine gives the* indigo colour reaction with ferric chloride, a test characteristic of cysteine, this suggests that the amine is probably a substituted cysteine. The structure of penicillamine was proved to be D-$\beta$$\beta$-dimethylcysteine by synthesis, *e.g.*,

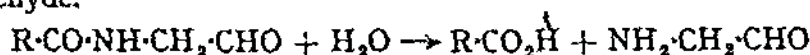
The racemic amine was resolved as follows: the amine was converted into the formyl derivative, which was then resolved by means of brucine. D-Penicillamine was obtained after removal of the formyl group by hydrolysis.

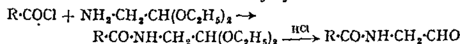
This was found to be identical with the natural penicillamine.

When treated with diazomethane, penicillin is converted into its methyl ester and this, on treatment with an aqueous solution of mercuric chloride, gives the methyl ester of penicillamine. Thus the carboxyl group in penicillamine is the carboxyl group in penicillin itself.

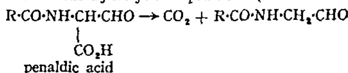
Penilloaldehyde. On vigorous hydrolysis, all the penilloaldehydes give a substituted acetic acid and aminoacetaldehyde. Thus the penilloaldehydes are acylated derivatives of aminoacetaldehyde.



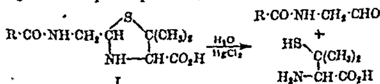
This structure has been confirmed by synthesis:



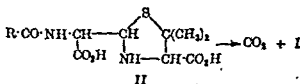
As pointed out above, the acid hydrolysis of penicillin gives penicillamine, penilloaldehyde and carbon dioxide. The formation of this molecule of carbon dioxide gave rise to the belief that it is formed by the ready decarboxylation of an unstable acid. Such an acid is a β-keto-acid, and so a possible explanation is that penilloaldehyde-carboxylic acid (penaldic acid) is formed as an intermediate in the hydrolysis of penicillin (see also below):



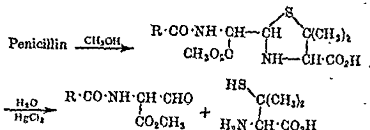
The problem now is: How are the two fragments, penicillamine and penilloaldehyde, combined in penicillin? The hydrolysis of penicillin with dilute alkali or with the enzyme penicillinase produces *penicilloic acid* (a dicarboxylic acid), which readily eliminates a molecule of carbon dioxide to form *penilloic acid*. This suggests that a carboxyl group is in the β-position with respect to a negative group (*cf.* above). Penilloic acid, on hydrolysis with aqueous mercuric chloride, gives penicillamine and penilloaldehyde. This hydrolysis is characteristic of compounds containing a thiazolidine ring (*cf.* §5b. XII). Thus penilloic acid could be I, since this structure would give the required products.



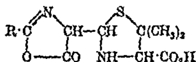
Hence, if I is penilloic acid, then penicilloic acid would be II.



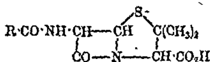
This structure (II) is supported by the fact that the treatment of penicillin with methanol gives methyl penicilloate which, on hydrolysis with aqueous mercuric chloride, gives methyl penaldate (see also above) and penicillamine.



On the basis of the foregoing evidence, two structures are possible for penicillin, *viz.*, III and IV.



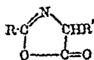
III
oxazolone structure



IV
β-lactam structure

It was not possible to decide between these two on chemical evidence alone, since penicillin readily undergoes molecular rearrangements, *e.g.*, on treatment with dilute acid, penicillin rearranges to penillic acid. It was therefore necessary to examine the molecule by physical methods (thereby leaving the molecule intact).

(i) Infra-red spectra studies showed the presence of two double bonds; these were exclusive of the C=O group in the carboxyl group in penicillin. The examination of the infra-red spectra of a number of oxazolones (these contain two double bonds, C=O and C=N) showed that this ring structure could not account for the absorption maxima obtained for penicillin. Thus structure III is untenable. On the other hand, it was found from an examination



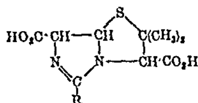
oxazolone

of the spectra of a number of amides that an amide structure could account for the spectrum of penicillin; thus IV is the probable structure of penicillin.

(ii) The X-ray analysis of the sodium, potassium and rubidium

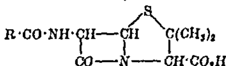
salts of benzylpenicillin showed the presence of a β-lactam ring; thus IV is the structure of penicillin.

Using this structure, we can now formulate the chemical reactions described above.

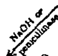


penillic acid

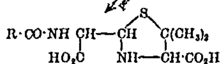




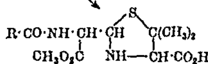
penicillin





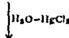


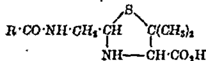
penicilloic acid



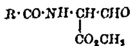
methyl penicilloate





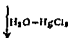


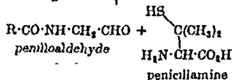
penilloic acid



methyl penaldate

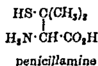
+





penilloaldehyde

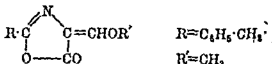
penicillamine



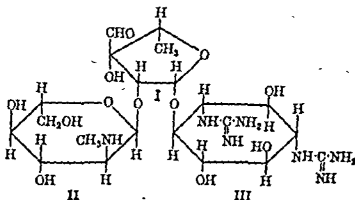
penicillamine

Penicillin has been synthesised by condensing synthetic D-penicillamine with a suitably substituted oxazolone containing a potential

aldehyde group, *e.g.*,

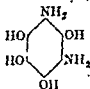


§7. **Streptomycin.** Streptomycin was isolated by Waksman *et al.* (1944) from cultures of *Streptomyces griseus*. This antibiotic is very effective in the treatment of tuberculosis, meningitis and pneumonia. Streptomycin is a solid with a lævorotation, and its structure has been shown to be composed of the three units streptose, I, *N*-methyl-L-glucosamine, II, and streptidine, III.



The following is a very brief account of the evidence that led to this structure for streptomycin. The molecular formula was shown to be $C_{21}H_{39}O_{12}N_7$. Three nitrogen atoms are strongly basic (the molecule forms a trihydrochloride), and on mild acid hydrolysis, streptomycin gives one molecule of streptidine, $C_8H_{18}O_4N_6$, and one molecule of streptobiosamine, $C_{13}H_{23}O_9N$ (Folkers *et al.*, 1945).

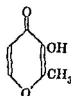
Streptidine (unit III), on oxidation with potassium permanganate, gave two molecules of guanidine (Peck *et al.*, 1946); thus two guanido groups are present in streptidine. Streptidine, on alkaline hydrolysis, gave **streptamine** and ammonia (Brink *et al.*,



1945). Streptamine was shown to be a diaminotetrahydroxy*cyclo*hexane, and examination of the oxidation products of dibenzoylstreptamine with periodic acid led to the suggestion that streptidine is 1;3-diguanido-2:4:5:6-tetrahydroxy*cyclo*hexane (Carter *et al.*, 1946). Streptidine has been synthesised from streptamine (Wolfrom *et al.*, 1948). Since streptidine is not optically active, the configuration of the molecule must be *meso*, with the two guanido groups *cis* (see unit III).

N-Methyl-L-glucosamine (unit II). When streptomycin is treated with methanolic hydrogen chloride (methanolysis), and then subjected to acid hydrolysis followed by acetylation, the penta-acetate of *N*-methyl-L-glucosamine is obtained; the parent compound is obtained by hydrolysis. The structure of *N*-methyl-L-glucosamine was confirmed by synthesis from L-arabinose (Kuehl *et al.*, 1946, 1947).

Streptose (unit I). The streptose fragment has not been isolated from streptomycin by degradation. It appears to be too unstable, but its structure was elucidated by various degradative experiments, *e.g.*, the alkaline hydrolysis of streptomycin gives

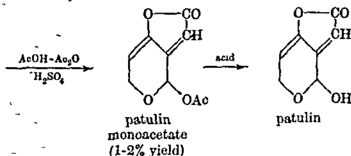


maltol

maltol (Schenck *et al.*, 1945), and this is produced by the conversion of a *furanose ring* into γ-pyrone.

Streptobiosamine (units I and II). Analytical work showed that this compound was a disaccharide, and from it was isolated *N*-methyl-L-glucosamine (see above). The formation of maltol and other analytical work led to the structure (I + II) for streptobiosamine, and then the points of attachment between streptobiosamine and streptidine were found, and so led to the structure given above for streptomycin (Kuehl *et al.*, 1947, 1948).

§7a. **Aureomycin** and **Terramycin.** Aureomycin was isolated from cultures of *Streptomyces aureofaciens*, and is used in the treatment of typhoid fever, etc. Terramycin was isolated from cultures of *Streptomyces rimosus*, and is very effective in the treatment of trachoma. The structures of these antibiotics are (Woodward *et al.*, 1952):

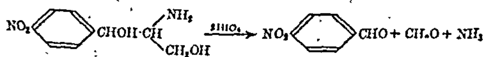


The monoacetate (obtained above) was shown to be identical with that obtained from patulin.

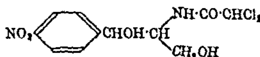
§9. **Chloramphenicol** (Chloromycetin). Chloramphenicol is a lævorotatory compound that is produced by *Streptomyces venezuelæ* (Carter *et al.*, 1948); it is very effective in the treatment of typhoid fever, etc.

The molecular formula of chloramphenicol is $C_{11}H_{12}O_5N_2Cl_2$, and its absorption spectrum is similar to that of nitrobenzene. The presence of a nitro-group was shown by the reduction of chloramphenicol with tin and hydrochloric acid, followed by diazotisation and then coupling to give an orange-red precipitate with 2-naphthol (Rebstock *et al.*, 1949). When catalytically reduced (palladium), chloramphenicol gives a product which has an absorption spectrum similar to that of *p*-toluidine, and the solution contains ionic chlorine. The hydrolysis of chloramphenicol with acid or alkali produces dichloroacetic acid and an optically active base, $C_9H_{12}O_4N_2$. This base was shown to contain a primary amino-group, and when treated with methyl dichloroacetate, the base reformed chloramphenicol (Rebstock *et al.*, 1949).

Chloramphenicol is converted into a diacetyl derivative on treatment with acetic anhydride in pyridine; the base obtained from chloramphenicol forms a triacetyl derivative on similar treatment. Thus chloramphenicol probably contains two hydroxyl groups. When the base is treated with periodic acid, two molecules of the latter are consumed with the formation of one molecule each of ammonia, formaldehyde and *p*-nitrobenzaldehyde. These products may be accounted for if the base is assumed to be 2-amino-1-*p*-nitrophenylpropane-1:3-diol (Rebstock *et al.*, 1949).



Thus chloramphenicol will be



This structure has been confirmed by synthesis, *e.g.*, that of Long *et al.* (1949).

$NO_2C_6H_4\cdot CO\cdot CH_3 \xrightarrow{Br_2} NO_2C_6H_4\cdot CO\cdot CH_2Br \xrightarrow[\text{(ii) } HCl-C_2H_5OH]{\text{(i) } (CH_2)_6N_4}$

$NO_2C_6H_4\cdot CO\cdot CH_2\cdot NH_2 \xrightarrow{(CH_3CO)_2O} NO_2C_6H_4\cdot CO\cdot CH_2\cdot NH\cdot CO\cdot CH_3$

$\xrightarrow[Na_2CO_3\ aq.]{CH_2O} NO_2C_6H_4\cdot CO\cdot CH(NH\cdot CO\cdot CH_3)(CH_2OH) \xrightarrow{[(CH_3)_2CHO]_3Al}$

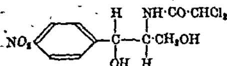
$NO_2C_6H_4\cdot CHOH\cdot CH(NH\cdot CO\cdot CH_3)(CH_2OH) \xrightarrow{HCl} NO_2C_6H_4\cdot CHOH\cdot CH(NH_2)(CH_2OH)$

$\xrightarrow[\text{(ii) } CHCl_2\cdot CO_2CH_3]{\text{(i) resolved}} NO_2C_6H_4\cdot CHOH\cdot CH(NH\cdot CO\cdot CHCl_2)(CH_2OH)$

(−)-chloramphenicol

This structure has also been confirmed by crystallographic studies (Dunitz, 1952).

Chloramphenicol and the base contain two asymmetric carbon atoms; thus there are two possible pairs of enantiomorphs. Comparison of the properties of the base with those of norephedrine and *nor-ψ*-ephedrine (§7. XIV) showed that the configuration of the base was similar to that of nor-ψ-ephedrine (Rebstock *et al.*, 1949). Thus chloramphenicol is D-(−)-*threo*-2-dichloroacetamido-1-*p*-nitrophenylpropane-1:3-diol.



It is interesting to note that chloramphenicol is the first natural

compound found to contain a nitro-group, the presence of the CCl_3 group is also most unusual.

Reading References

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley. Vol III (1953) (i) Ch. 5. Some Aspects of Chemotherapy. (ii) Ch 6 Antibiotics.

Raiziss and Gavron, *Organic Arsenical Compounds*, Chemical Catalog Co. (1923).

Work anc

Northey,

Northey,

Derivatives, *Chem Reviews*, 1940, **27**, 85

Ann. Reports (Chem. Soc), 1950, **47**, 285 Antibiotics.

Haynes, Physiologically Active Unsaturated Lactones, *Quart. Reviews (Chem. Soc)*, 1948, **2**, 46.

Cook, The Chemistry of the Penicillins, *Quart Reviews (Chem. Soc)*, 1948, **2**, 203.

The Chemistry of Penicillin, Princeton University Press (1949).

Antibiotics, Oxford Press (2 volumes; 1949).

Robinson, *Antibiotics*, Pitman (1953).

Waksman (Ed.), *Streptomycin*, Williams and Wilkins Co. (1949).

Lemieux and Wolfrom, The Chemistry of Streptomycin, *Advances in Carbohydrate Chemistry*, Academic Press, 1948, **3**, 337.

Brink and Folkers, Some Aspects of Streptomycin and Other *Streptomyces* Antibiotics, *Advances in Enzymology*, Interscience Publishers, 1950, **10**, 145.

Birkinshaw, The Chemistry and Biochemistry of Streptomycin and Related Compounds, *J. Pharm. Pharmacol.*, 1951, **3**, 529.

Rebstock *et al.*, Chloramphenicol, *J. Amer. Chem. Soc.*, 1949, **71**, 2458, 2463.

Long *et al.*, Chloramphenicol, *J. Amer. Chem. Soc.*, 1949, **71**, 2469, 2473.

Rose, A Chemotherapeutic Search in Retrospect, *J.C.S.*, **1951**, 2770.

Barber, Chance and Design in the Search for New Drugs, *Chem. and Ind.*, **1955**, 1460.

Burger, Rational Approaches to Drug Structure, *J. Chem. Educ.*, 1956, **33**, 362.

Bracken, *The Chemistry of Micro-Organisms*, Pitman (1955).

CHAPTER XIX

HÆMOGLOBIN, CHLOROPHYLL AND PHTHALOCYANINES

§1. **Introduction.** Two of the most important compounds of the natural porphyrins are hæmoglobin and chlorophyll. The bile pigments, which are formed mainly in the liver, are degradation products of hæmoglobin. Hæmoglobin and chlorophyll act as catalysts (biological) in many biological processes.

HÆMOGLOBIN

§2. **Degradation products of hæmoglobin.** Hæmoglobin occurs in the red blood corpuscles of mammals. It is a chromoprotein (§7 B. XIII), the protein part being *globin* (94 per cent.), and the prosthetic group being *hæm* (6 per cent.). The composition of hæmoglobin varies slightly, depending on the animal species from which it is isolated; the variation occurs only in the globin part of the molecule. It is interesting to note that hæmoglobin was the first protein to be obtained in a crystalline form.

In the animal body, hæmoglobin readily combines with oxygen to form *oxyhæmoglobin*, and this, when treated with glacial acetic acid, forms *hæmatin*, $C_{34}H_{32}O_4N_4Fe^{III}\cdot OH$. The chloride of hæmatin is known as *hæmin*; its molecular formula is $C_{34}H_{32}O_4N_4Fe^{III}Cl$ (the chlorine is ionised, and the iron atom is in the ferric state). Hæmin may be prepared by warming blood with acetic acid and sodium chloride (Teichmann, 1853). The iron can be removed from hæmin, and replaced. The iron-free compounds are known as *porphyrins*, and the iron-containing compounds as *hæms*; the nature of the porphyrin depends on the conditions which are used to remove the iron atom from hæmin. When hæmin is reduced with sodium hyposulphite, the base *hæm* is produced in which the atom of iron is in the bivalent state; the molecular formula of hæm is $C_{34}H_{32}O_4N_4Fe$.

Since hæmin forms a diester with methanol, the molecule therefore contains two carboxyl groups. Also, since hæmin absorbs two molecules of hydrogen when catalytically reduced (palladium), two ethylenic double bonds are thus probably present in the molecule.

When subjected to vigorous reduction with hydriodic acid and phosphonium iodide or hydriodic acid and acetic acid, hæmin is degraded into the four pyrrole derivatives opsopyrrole, I, hæmopyrrole, II, cryptopyrrole, III, and phyllopyrrole, IV.

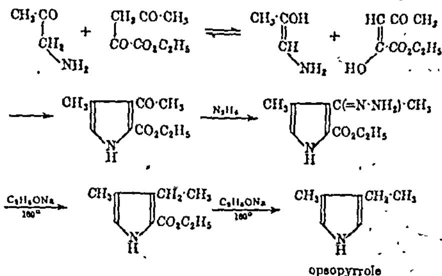
CH_3, C_2H_5; N–H	CH_3, C_2H_5, CH_3; N–H	CH_3, C_2H_5, CH_3; N–H	CH_3, C_2H_5, CH_3, CH_3; N–H
opsopyrrole I	hæmopyrrole II	cryptopyrrole III	phyllopyrrole IV

All four compounds have been synthesised by means of the Knorr pyrrole synthesis (1884, 1886); this is the condensation between an α-aminoketone and a ketone containing an active methylene group, *i.e.*, a compound containing the group $—CH_2·CO—$. The mechanism of the reaction is not known; possibly the enol forms are involved, and so we may write the general reaction as follows:

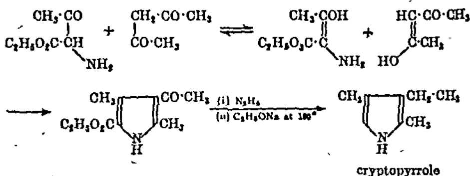
$$R·CO·CH(R')·NH_2 + H_2C(R'')·CO·R''' \rightleftharpoons R·C(OH){=}C(R')·NH_2 + HC(R''){=}C(OH)·R'''$$

$$\longrightarrow \text{pyrrole ring } R·C{=}C·R'' ,\ R'·C{=}C·R''' ,\ N–H + 2H_2O$$

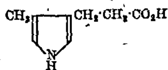
A detailed study of this reaction has shown that the yields depend on the nature of R, R', R'' and R'''; when R' and R'' are acyl or carbalkoxyl groups, the yields are usually very good. As examples of the Knorr synthesis, let us consider the preparation of opsopyrrole (I) and cryptopyrrole (III). Opsopyrrole may be synthesised by condensing aminoacetone with ethyl 2:4-diketopentanoate, and then subjecting the product to the Wolff-Kishner reduction, *i.e.*, first converting the product into the hydrazone and then heating the latter with sodium ethoxide at 160°. By this means a keto group is converted into a methylene group (see also Vol. I). By using an excess of sodium ethoxide, decarboxylation is also effected at the same time.



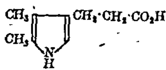
Cryptopyrrole may be prepared in a similar manner, starting from ethyl α-aminoacetoacetate and acetylacetone (penta-2:4-dione).



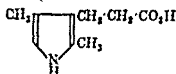
When reduced with tin and hydrochloric acid, hæmin is again degraded into four pyrrole derivatives, but in this case the products are all carboxylic acids in which each of the four pyrroles I–IV contains a carboxyl group attached to the ethyl group:



opsopyrrole-carboxylic acid
V

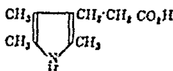


hæmopyrrole-carboxylic acid
VI



cryptopyrrole-carboxylic acid

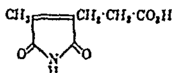
VII



phyllopyrrole-carboxylic acid

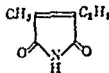
VIII

When oxidised with chromic acid, hæmin gives two molecules of *hæmatinic acid* (IX). On the other hand, *mesoporphyrin* (see below) gives, on oxidation, two molecules of ethylmethylmaleimide (X).



hæmatinic acid

IX



ethylmethylmaleimide

X

The treatment of hæmin with iron dust and formic acid results in the removal of the iron atom and the formation of *protoporphyrin*, $C_{34}H_{34}O_4N_4$. The iron atom is also removed from hæmin by the action of hydrobromic acid in acetic acid, but in this case the product is *hæmatoporphyrin*, $C_{34}H_{38}O_6N_4$. If, however, hæmin is treated with hydriodic acid in acetic acid, the iron atom is again removed and *mesoporphyrin*, $C_{34}H_{38}O_4N_4$, is obtained.

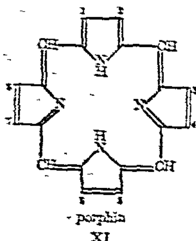
Finally, when porphyrins containing two carboxyl groups are decarboxylated, the products obtained (after reduction, if necessary) are known as *ætioporphyrins*, *e.g.*, when protoporphyrin is decarboxylated, and the product then reduced, the final product is *ætioporphyrin*, $C_{32}H_{38}N_4$, which is also a degradation product of chlorophyll. Thus hæmin and chlorophyll are closely related chemically.

The table on page 780 summarises the reactions that have been discussed.

Compound	Reaction	Products
Hæmoglobin	Atmospheric oxidation	Oxyhæmoglobin
Oxyhæmoglobin	$CH_3 \cdot CO_2H$	Hæmatin
Oxyhæmoglobin	$CH_3 \cdot CO_2H + NaCl$	Hæmin
Hæmin	$Na_2S_2O_4$	Hæm
Hæmin	$HI + PH_4I$	[illegible]
Hæmin	Sn—HCl	[illegible] carboxylic acids
Hæmin	CrO_3—H_2SO_4	Hæmatinic acid
Mesoporphyrin	CrO_3—H_2SO_4	Ethylmethylmaleimide
Hæmin	Fe—$H \cdot CO_2H$	Protoporphyrin
Hæmin	HBr—$CH_3 \cdot CO_2H$	Hæmatoporphyrin
Hæmin	HI—$CH_3 \cdot CO_2H$	Mesoporphyrin
Porphyrin	Decarboxylation (and then reduction, if necessary)	Ætioporphyrins

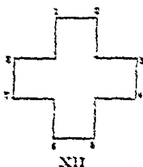
are linked at the α-positions *via* one carbon atom. The isolation of two molecules of IX suggests the presence of two propionic acid residues each in the β-position of two pyrrole nuclei (this would also account for the two carboxyl groups present in hæmin). The appearance of ethyl groups in I–IV on the reduction of hæmin could be explained by the presence of two vinyl groups in the β-position of two pyrrole nuclei (hæmin contains two ethylenic double bonds). A possible structure for hæmin is thus a ring structure containing four pyrrole nuclei linked at the α-positions *via* one carbon atom, with four β-positions occupied by methyl groups, two β′-positions by vinyl groups, and the remaining two β′-positions by propionic acid residues. Küster (1912) was the first to propose that the four pyrrole nuclei formed a cyclic structure, and this has been proved correct by synthetic work; the porphyrins so obtained had the same absorption spectra as the natural porphyrins. At the same time, this synthetic work established the nature and the positions of the substituent groups.

The parent substance of all the compounds mentioned above is *porphin* (XI), and this may conveniently be written as XII (H. Fischer). In this porphin molecule there is an eighteen-membered ring containing a complete arrangement of conjugated double bonds. Thus many resonating structures contribute to this molecule, and consequently its stability will be great; this is observed in practice, *e.g.*, the molecule has a very large heat of combustion. Also, the resonance gives rise to the colour in porphin derivatives (see Ch. XXXI, Vol. I); porphin itself does not occur naturally. It has been shown, by analogy with the X-ray data on phthalocyanines



porphin

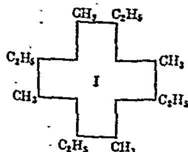
XI

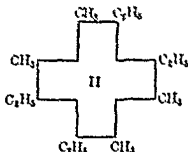


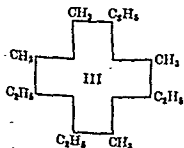
XII

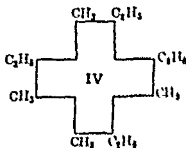
(59), that the porphin molecule is planar, and this planar structure is also in agreement with magnetic measurements.

The ætioporphyrins, $C_{32}H_{38}N_4$, are derivatives of porphin in which the 3- and 4-positions of each pyrrole nucleus are substituted by methyl and ethyl groups. Four isomers are possible, and these are known as ætioporphyrin I, II, III and IV, respectively.







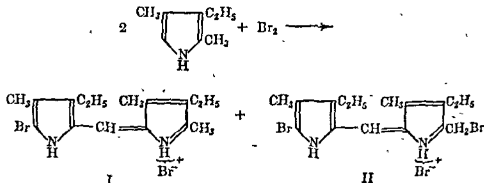


ætioporphyrins

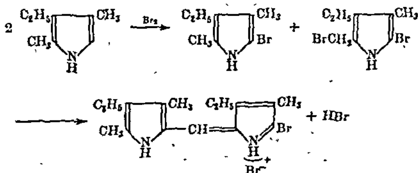
All four ætioporphyrins have been synthesised; the degradation of hæmin gives ætioporphyrin III.

§3. Synthesis of the porphyrins. The first step in the synthesis of porphyrins is the synthesis of the dipyrrylmethenes.

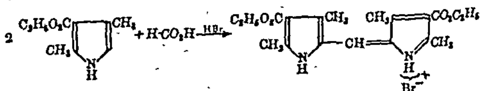
(i) *Dipyrrylmethenes may be prepared by the bromination of a 2-methylpyrrole in which position 5 is vacant* (H. Fischer, 1915); at least two products are obtained, *e.g.*, cryptopyrrole gives compounds I and II.



According to Corwin *et al.* (1944), the mechanism of this reaction is:



(ii) When pyrroles, in which the 5-position is vacant, are coupled by means of formic acid in the presence of hydrobromic acid, dipyrrylmethenes are produced (H. Fischer *et al.*, 1922); *e.g.*,



(iii) Piloty *et al.* (1914) showed that dipyrrylmethanes may be oxidised to the corresponding methenes by means of bromine, *e.g.*,

$$\text{(dipyrrylmethane: } C_2H_5O_2C, CH_3, CH_3 \text{ / } -CH_2- \text{ / } CH_3, CO_2C_2H_5, CH_3\text{)} \xrightarrow{Br_2} \text{(dipyrrylmethene hydrobromide: } -CH= \text{, } \overset{+}{N}H\ Br^-\text{)}$$

H. Fischer *et al.* (1923) modified the above procedure as follows. A dipyrrylmethane containing carbethoxyl groups was first prepared, this then hydrolysed and then treated with bromine in acetic acid. In this way the methane derivative is oxidised to the methene compound, but at the same time the carboxyl groups in position 5:5′ are replaced by bromine atoms, *e.g.*,

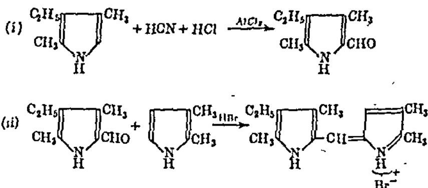
$$\text{(pyrrole: } CH_3, CO_2C_2H_5, C_2H_5O_2C, CH_3, NH\text{)} \xrightarrow{Br_2} \text{(pyrrole: } CH_3, CO_2C_2H_5, C_2H_5O_2C, CH_2Br, NH\text{)} \xrightarrow{\text{boil in water}}$$

$$\left[\text{(pyrrole: } CH_3, CO_2C_2H_5, C_2H_5O_2C, CH_2OH, NH\text{)}\right] \longrightarrow \text{(dipyrrylmethane: } CH_3, C_2H_5O_2C, CO_2C_2H_5 \text{ / } -CH_2- \text{ / } CO_2C_2H_5, CH_3, CO_2C_2H_5\text{)} + H_2O + CH_2O$$

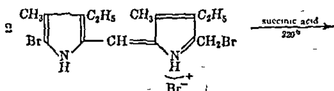
$$\xrightarrow[\text{(ii) } Br_2-CH_3CO_2H]{\text{(i) NaOH}} \text{(dipyrrylmethene hydrobromide: } CH_3, Br, CO_2H \text{ / } -CH= \text{ / } CO_2H, CH_3, Br, \overset{+}{N}H\ Br^-\text{)}$$

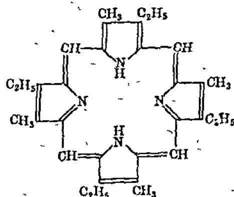
(iv) The foregoing methods (except i) lead to the formation of *symmetrical* dipyrrylmethenes. The preparation of *unsymmetrical*

dipyrrylmethenes is best carried out as follows, using the Gattermann aldehyde synthesis (Piloty *et al.*, 1912, 1914; H. Fischer *et al.*, 1926); *e.g.*,



The dipyrrylmethenes are coloured solids. H. Fischer *et al.* (1926)





ætioporphyrin I

of methanol and pyridine (Rothemund, 1936, 1939; Calvin *et al.*, 1943).

4 pyrrole + 4 CH_2O ⟶ [cyclic tetrapyrrylmethane: four pyrrole rings (NH, NH, HN, NH) linked by CH_2 bridges]

⟶ [porphin: four pyrrole rings linked by CH bridges]

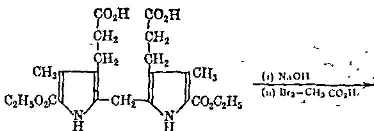
porphin

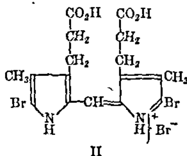
It should be noted that the two imino hydrogen atoms are replaced by the iron atom in the hæms, and the iron atom is covalently bound.

§4. **Synthesis of hæmin** (H. Fischer *et al.*, 1929).

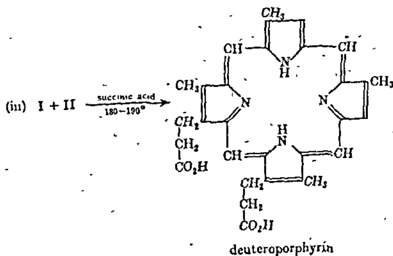
(i) 3,4-dimethylpyrrole + 2-formyl-3,5-dimethylpyrrole (OCH, CH_3, CH_3) $\xrightarrow{HBr}$ dipyrrylmethene hydrobromide (CH_3, CH_3 ... CH= ... CH_3, CH_3, $\overset{+}{N}H\}Br^-$) **I**

(ii) pyrrole (CH_3, $CH_2{\cdot}CH_2{\cdot}CO_2H$, $C_2H_5O_2C$, CH_3, NH) $\xrightarrow{Br_2}$ pyrrole (CH_3, $CH_2{\cdot}CH_2{\cdot}CO_2H$, $C_2H_5O_2C$, CH_2Br, NH) $\xrightarrow{\text{boil in water}}$

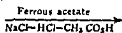


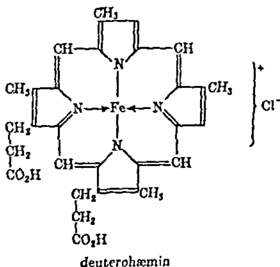


(iii) I + II

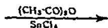


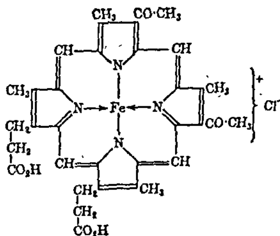
deuteroporphyrin



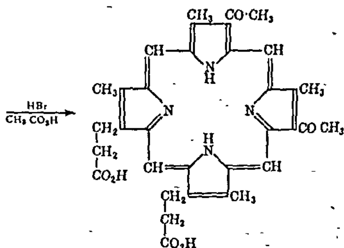


deuterohæmin

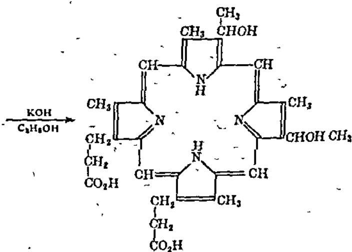




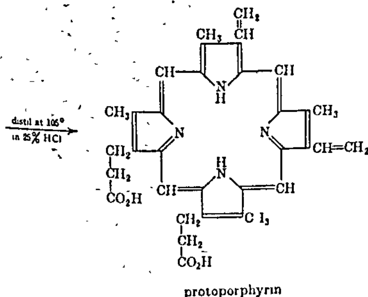
diacetyldeuterohæmin



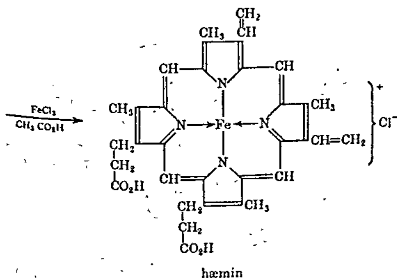
diacetyldeuteroporphyrin



hæmatoporphyrin



protoporphyrin



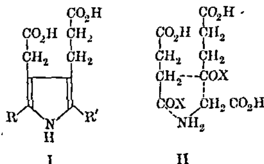
hæmin

It should be noted that the introduction of the iron atom into deuteroporphyrin to give deuterohæmin renders the pyrrole nuclei more reactive.

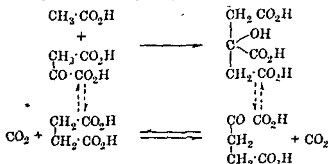
§4a. ***Biosynthesis of porphyrin.*** The progress made in this field is one of the outstanding examples of the use of isotopes. Tracer syntheses *in vivo* and *in vitro* and degradation methods have established the origin of all the carbon and nitrogen atoms in protoporphyrin (of hæm), and have also established the nature of the pyrrole precursors. These results are the outcome of a large volume of work, but in the following account only a few experiments have been mentioned. These indicate, to some extent, the lines of research pursued.

Bloch *et al.* (1945), using acetic acid labelled with deutero atoms, showed that deuterohæmin was produced. Thus at least the methyl carbon of acetic acid is involved in the biosynthesis of hæm. Then Shemin *et al.* (1950) and Neuberger *et al.* (1950) carried out experiments with $^{14}CH_3{\cdot}CO_2H$ and $CH_3{\cdot}^{14}CO_2H$, and showed that *both* carbon atoms of acetate participate in the synthesis of *hæm. The latter authors also showed that with* $^{14}CH_3{\cdot}CO_2H$, about half of the radioactive tracer atom appeared in the two pyrrole nuclei carrying the vinyl radicals, and the other half in the two pyrrole nuclei carrying the propionic acid residues. When, however, $CH_3{\cdot}^{14}CO_2H$ was used as the precursor, then about 20 per cent. of the tracer atom appeared in the vinyl pyrrole nuclei and 80 per cent. in the propionic acid pyrrole nuclei. In neither case of the labelled acetates was there any significant radioactivity in the methine carbon of the hæm. Thus the carbons of the methine bridges do not originate from acetate.

Shemin *et al.* (1945, 1946) carried out experiments with [^{15}N] glycine, and showed that all the nitrogen atoms in hæm are derived from this glycine. Shemin *et al.* (1950) also used $CH_2(NH_2){\cdot}^{14}CO_2H$, and showed that the carboxyl group of glycine is *not incorporated* into protoporphyrin. On the other hand, Altman *et al.* (1948), using $^{14}CH_2{\cdot}NH_2{\cdot}CO_2H$, showed that the α-carbon atom of glycine is used in the protoporphyrin synthesis. This was confirmed by Shemin *et al.* (1950) who used $^{14}CH_2{\cdot}^{15}NH_2{\cdot}CO_2H$ and showed that for each nitrogen used for hæm synthesis, two α-carbon atoms of glycine were also incorporated into the molecule. Similar results were obtained by Neuberger *et al.* (1950) who also showed that the α-carbon atom of glycine is used in the formation of the methine bridge. Thus all the carbon atoms of protoporphyrin, except eight derived from the α-carbon of glycine, originate from acetate. Furthermore, a detailed study of the degradation products of the labelled protoporphyrins showed that it was very probable that the two sides of the pyrrole nuclei were synthesised from identical intermediates. It also seemed very reasonable that a *common* pyrrole of the type I was formed first. Also, consideration of the

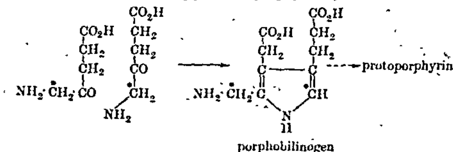


distribution of the radioactivity of the carbon atoms of the propionic acid residue and the (pyrrole) nuclear carbon to which it was attached led to the suggestion that succinic acid was a precursor, and that two molecules of this, on condensation with one molecule of glycine, could form the common pyrrole (I). The tracer distribution of the labelled succinic acid could arise by acetate entering the *Krebs cycle* (§18. XIII). Two molecules of



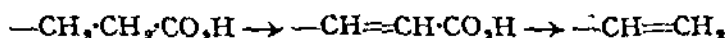
"active" succinate (succinyl-coenzyme A) and one of glycine then forms the common precursor (see II). Shemin *et al.* (1952) tested this succinic acid hypothesis by using $^{14}CO_2H{\cdot}CH_2{\cdot}CH_2{\cdot}^{14}CO_2H$ and $CO_2H{\cdot}^{14}CH_2{\cdot}^{14}CH_2{\cdot}CO_2H$, and showed that hæm contained the labelled carbon.

In 1952, Westall isolated porphobilinogen from the urine of humans suffering from acute porphyria. Based on this, Shemin *et al.* (1953) now proposed that δ-aminolævulic acid can replace "active" succinate and glycine in porphyrin synthesis:



This pyrrole synthesis is supported by various experiments, *e.g.*, Shemin *et al.* (1954) used [δ^{14}C]δ-aminolævulic acid as precursor, and showed that half of the radioactivity is equally distributed among the four pyrrole nuclei and the other half is in the methine-bridge carbons. This distribution is in agreement with the equation given. Furthermore, Falk *et al.* (1953) have shown that porphobilinogen is the common precursor in porphyrin synthesis.

The problem of the conversion of porphobilinogen into protoporphyrin has still to be elucidated. Decarboxylation of the acetic acid radicals would produce the methyl radicals (in protoporphyrin). The conversion of the propionic acid residues into vinyl radicals takes place by a series of steps; a possibility is:

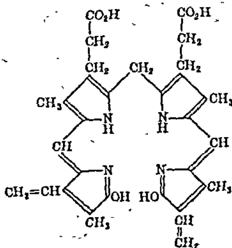


$$—CH_2·CH_2·CO_2H \rightarrow —CH{=}CH·CO_2H \rightarrow —CH{=}CH_2$$

other products, bilirubic acid, $C_{17}H_{24}O_4N_2$, and neobilirubic acid, $C_{14}H_{22}O_3N_2$. Finally, the reduction of bilirubic acid gives cryptopyrrolecarboxylic acid as the main product, and the reduction of

to show that bilirubin is a degradation product of hæmoglobin.

Since the absorption spectrum of bilirubin is not like that of a porphyrin, it is assumed that bilirubin has an *open-chain* structure. Further degradative and synthetic work has shown that bilirubin probably has the following structure



bilirubin

CHLOROPHYLL

§6. Introduction. Chlorophyll is the green colouring matter of leaves and green stems, and its presence is essential for photosynthesis. Photosynthesis is the process in which light energy is used by plants to synthesise carbohydrates, proteins and fats. In green plants it is the chlorophyll which absorbs the light energy.

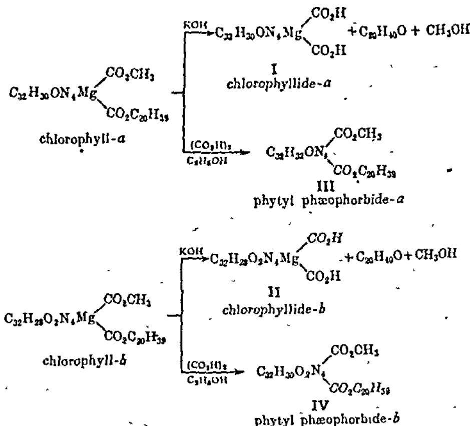
The name *chlorophyll* was given to the green pigment in leaves by Pelletier and Caventou (1818). There the matter rested until 1864, when Stokes showed, from spectroscopic evidence, that chlorophyll was a mixture. This paper apparently did not attract much attention, and it was not until Willstätter entered the field that any progress in the chemistry of chlorophyll was made.

When dried leaves are powdered and then digested with ethanol, a "crystalline" chlorophyll is obtained after concentration of the solvent. If, however, ether or aqueous acetone is used instead of ethanol, then the product is "amorphous" chlorophyll (Willstätter *et al.*, 1908). The extraction of chlorophyll is also accompanied by the extraction of two other pigments, carotene and xanthophyll (see Ch. IX). Willstätter *et al.* (1910) then showed that "crystalline" chlorophyll was produced during the extraction of chlorophyll by means of ethanol, a molecule of phytyl alcohol being replaced by ethanol under the influence of an enzyme, chlorophyllase (which is present in leaves). Nettle leaves are the main source for the extraction of chlorophyll on a large scale.

Willstätter *et al.* (1911) originally gave chlorophyll the molecular formula $C_{55}H_{72}O_6N_4Mg$, but in 1912 Willstätter *et al.* showed that chlorophyll, obtained from a wide variety of sources, was a mixture of two compounds, chlorophyll-*a* and chlorophyll-*b*. The separation was effected by shaking a light petrol solution of chlorophyll with aqueous methanol; chlorophyll-*a* remains in the light petrol, and chlorophyll-*b* passes into the aqueous methanol. Chlorophyll-*a* is a bluish-black solid, giving a green solution in organic solvents; chlorophyll-*b* is a dark green solid, also giving a green solution in organic solvents. The two components occur in proportions of approximately 3 of *a* to 1 of *b* in natural chlorophyll. Winterstein *et al.* (1933) have separated the two chlorophylls by means of chromatography (on sucrose as adsorbent).

The molecular formulae that have been assigned to chlorophyll-*a* and chlorophyll-*b* are $C_{55}H_{72}O_5N_4Mg$ and $C_{55}H_{70}O_6N_4Mg$, respectively (Willstätter, 1913); the two compounds have different absorption spectra (*cf.* Stokes, above). The hydrolysis of both

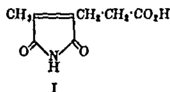
chlorophylls with cold dilute potassium hydroxide solution gives one molecule of phytol, $C_{20}H_{40}O$ (see §30. VIII), one molecule of methanol, and one molecule of chlorophyllide-*a* (chlorophyllin-*a*), I, or chlorophyllide-*b* (chlorophyllin-*b*), II. Thus the chlorophylls are di-esters. When either chlorophyll is heated with an ethanolic solution of hydrated oxalic acid, the magnesium atom is replaced by two hydrogen atoms to produce phytyl phæophorbide-*a* (III) or *b* (IV; these phytyl phæophorbides are also known as phæophytins *a* and *b*, and "crystalline" chlorophyll is ethyl chlorophyllide). The foregoing reactions may be formulated as follows:



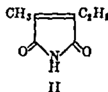
§6a. Nomenclature of the chlorophyll degradation products. Porphyrins are substituted porphins (see §2). *Phyllins, phyllides* and chlorophylls contain magnesium, whereas phorbins, phorbides and phytins are magnesium-free compounds, the magnesium atom having been removed and replaced by two hydrogen atoms. 7:8-Dihydroporphin is the nucleus of the *chlorin* series of compounds (tricarboxylic derivatives) which are derived from chlorophyll-*a*;

rhodins are the corresponding compounds derived from chlorophyll-*b*. The introduction of the extra ring—two methylene groups across the 6:γ-positions (see §7)—gives rise to the *phorbins*. The prefix phæo designates those compounds which have the same substituents that occur in chlorophyll. Chlorin itself is dihydroporphin, and the natural *red* porphyrin pigments are derivatives of porphin, whereas the *green* chlorophylls and their derivatives are derivatives of chlorin. Furthermore, examination of formulæ XIV and XV (in §7) shows that there is still complete conjugation in chlorin as in porphin (formula XI, §2). Chlorin has been synthesised by Linstead *et al.* (1955), and has been dehydrogenated to porphin.

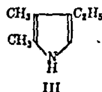
§7. Structure of chlorophyll-*a*. When phytyl phæophorbide-*a* is hydrolysed with boiling methanolic potassium hydroxide (30 seconds), the product is chlorin-*e*. This is a tricarboxylic acid (*e.g.*, it forms a trimethyl ester), and its molecular formula may thus be written as $C_{31}H_{33}N_4(CO_2H)_3$. Chlorin-*e*, on oxidation with chromic acid or with Caro's acid, gives hæmatinic acid, I, and ethylmethylmaleimide, II (Willstätter *et al.*, 1910). When



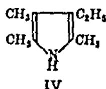
I



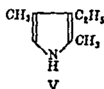
II



III



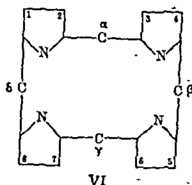
IV



V

chlorin-*e* is reduced with hydriodic acid in acetic acid, hæmopyrrole, III, and phyllopyrrole, IV, are produced (Willstätter *et al.*, 1911). When phylloporphyrin (see below) is reduced under the same conditions, the products are now III, IV, and cryptopyrrole, V. From these results it is reasonable to infer that chlorophyll-*a* contains four pyrrole nuclei, each probably having a methyl group in the β-position (see II–V). It is also reasonable to suppose that at least one pyrrole nucleus contains a propionic acid residue in the β'-position (see I). It also appears likely that a vinyl group is present in the molecule (this would account for the presence of an ethyl group on reduction; at the same time, the presence of an ethyl group, as such, is not excluded). Furthermore, the isolation of I

and II on oxidation (giving oxidation at the α-position), and of III and IV on reduction (the appearance of a methyl group at the α-position), can be interpreted as meaning that the four pyrrole nuclei are joined to each other at their α-positions *via* one carbon

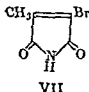


VI

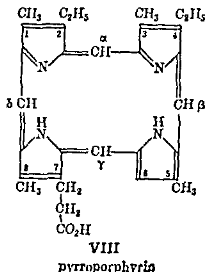
atom (*cf.* §2). Thus a possible skeleton structure for chlorin-*e* could be a cyclic one, VI; the positions of the various substituent groups cannot be assigned on the evidence obtained so far, *e.g.*, a methyl group at 1 and a propionic acid residue at 2 would produce the same oxidation product I had the positions of the *two* groups been interchanged in VI. It is also necessary to fit a second carboxyl group into this structure (VI), since chlorophyll-*a* forms chlorophyllide-*a* on hydrolysis (the latter compound contains *two* carboxyl groups). Furthermore, since chlorophyllide-*a*, on further hydrolysis, forms chlorin-*e*, a tricarboxylic acid, some group *must* be present which can give rise to this third carboxyl group. Such a group could be a lactone; it must be *cyclic* since no carbon atoms are lost after the hydrolysis.

By the further degradation of chlorin-*e*, *e.g.*, heating in a sealed tube with ethanolic potassium hydroxide, various porphyrins are obtained. Three of these are pyrroporphyrin, rhodoporphyrin and phylloporphyrin.

Pyrroporphyrin, $C_{30}H_{32}N_4{\cdot}CO_2H$, has an absorption spectrum closely resembling that of mesoporphyrin (see §2); this agrees with the tentative skeleton structure VI proposed for chlorin-*e*. Pyrroporphyrin, on bromination followed by oxidation with chromic acid, gives bromocitraconimide, VII, as one of the products (Treibs *et al.*, 1928). It therefore follows that at least one of the pyrrole nuclei in pyrroporphyrin has a free *β*-position available for bromination. Synthetic work then showed that pyrroporphyrin has structure VIII (H. Fischer *et al.*, 1929, 1930, 1933); thus the positions of the four methyl groups and the position of the propionic acid group are now established.



VII

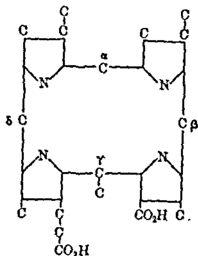


VIII
pyrroporphyrin

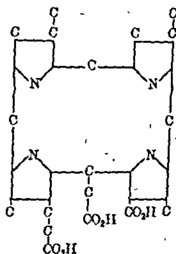
Rhodoporphyrin, $C_{30}H_{32}N_4(CO_2H)_2$, on heating with sodium ethoxide, readily loses *one* carboxyl group to form pyrroporphyrin (VIII). From a detailed study of the hæmin series, it was observed that a carboxyl group in a *side-chain* of a pyrrole nucleus was difficult to remove. Hence it is probable that the carboxyl group *lost from rhodoporphyrin is attached directly to a pyrrole nucleus.* The only position for this carboxyl group is at 6 (see structure VIII); elimination of the carboxyl group from rhodoporphyrin would then give one pyrrole nucleus with a free β-position (6), *i.e.*, pyrroporphyrin. Furthermore, comparison of the absorption spectra of rhodoporphyrin with compounds of known structure showed that the two carboxyl groups are in positions 6 and 7 (the latter is the propionic acid residue), and this was confirmed by the synthesis of rhodoporphyrin.

Phylloporphyrin, $C_{31}H_{35}N_4{\cdot}CO_2H$, *contains one* CH_2 *group more* than pyrroporphyrin, and may be converted into the latter by heating with sodium ethoxide. It therefore follows that the alkyl groups in both compounds occupy similar positions. Synthetic work then showed that phylloporphyrin contains a methyl group attached to the γ-methyne carbon atom (H. Fischer *et al.*, 1930, 1933).

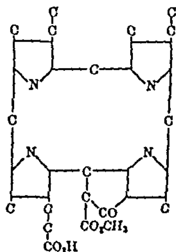
Consideration of the information obtained from the structures of the porphyrins described above shows that the skeleton structure IX is present in chlorin-*e*. Now chlorin-*e* contains three carboxyl groups and one more carbon atom than the structure shown in IX.



IX



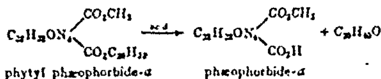
X



XI

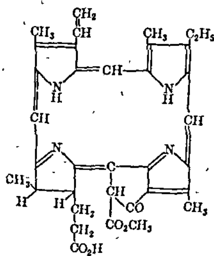
The formation of a methyl group (at the γ carbon atom) could be explained by assuming a carboxyl group is attached as shown in structure X.

When phytyl phæophorbide-*a* (III, §6) is hydrolysed with acid, the phytyl group is removed to form phæophorbide-*a*.

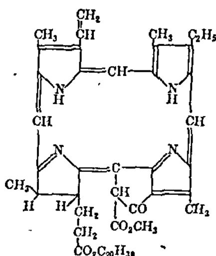


When phæophorbide-*a* is treated with hydriodic acid in acetic acid and followed by atmospheric oxidation, the product is phæoporphyrin-a_5. This, on further treatment with hydriodic acid in acetic acid, forms phylloerythrin, $C_{33}H_{34}O_3N_4$, by loss of the carbomethoxyl group; phylloerythrin has the same absorption spectrum as that of the porphyrins, and so the porphin structure is still present. Now both phæophorbide-*a* and phylloerythrin contain a keto group (as is shown by the formation of an oxime, etc.), and so when the carbomethoxyl group is hydrolysed, the elimination of carbon dioxide can be expected if the keto group is in the β-position with respect to the carboxyl group (produced on hydrolysis). Furthermore, the hydrolysis of phæophorbide-*a* with methanolic potassium hydroxide gives chlorin-*e*. In this reaction, apart from the hydrolysis of the carbomethoxyl group, the keto group is lost and a carboxyl group is introduced *without the loss of any carbon atoms*. This may be explained by assuming that this carboxyl group (the third one in chlorin-*e*) is produced by the fission of a *cyclic ketone*, and not from a lactone as suggested previously (see above). Thus a possible skeleton structure for phæophorbide-*a* is XI; if the ketone ring is opened, then the formation of X can be expected. Also, the hydrolysis of XI would produce a β-keto-acid, which can be expected to lose carbon dioxide readily to form phylloerythrin.

Phæophorbide-*a* can be reduced catalytically to its dihydro-derivative in which the keto group remains intact. This suggests



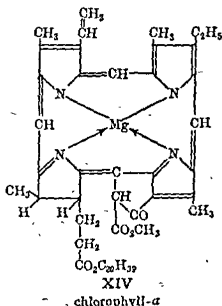
XII
phæophorbide-*a*



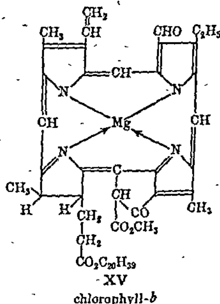
XIII
phytyl phæophorbide-*a*

the presence of a readily reducible double bond. Oxidation experiments on phæophorbide-*a* and dihydrophæophorbide-*a* showed the presence of one vinyl group in the former. Furthermore, the existence of a vinyl group in the ester of chlorin-*e* was shown by the reaction with diazoacetic ester to give a *cyclo*propane derivative, which was isolated by the oxidation of the addition product (H. Fischer *et al.*, 1935; *cf.* §2a. XII). Thus one of the ethyl groups (see pyrroporphyrin, VIII) must have been a vinyl group before reduction. Further degradative and synthetic work by H. Fischer *et al.* (1934–1936) showed that phæophorbide-*a* is XII and that phytyl phæophorbide-*a* is XIII.

The replacement of the two imino hydrogen atoms in XIII by a magnesium atom would therefore give chlorophyll-*a*; this is XIV. Chlorophyll-*b* has been assigned structure XV.



XIV
chlorophyll-*a*



XV
chlorophyll-*b*

PHTHALOCYANINES

§8. Preparation of the phthalocyanines. Phthalocyanines are a very important class of organic dyes and pigments; they are coloured blue to green. They were discovered by accident at the works of Scottish Dyes Ltd. in 1928. It was there observed that some lots of phthalimide, manufactured by the action of ammonia on molten phthalic anhydride in an iron vessel, were

contaminated with a blue pigment. The structure and method of formation of this compound were established by Linstead and his co-workers (1934).

The phthalocyanines form metallic complexes with many metals, and the colour depends on the nature of the metal (copper, magnesium, lead, etc.); greener shades are obtained by direct chlorination or bromination. The metal phthalocyanines are insoluble in water, and are used as pigments. They are made water-soluble by sulphonation, and these soluble salts are used as dyes.

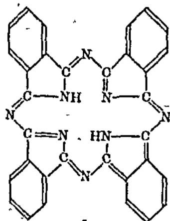
Metal phthalocyanines may be prepared as follows:

(i) By passing ammonia into molten phthalic anhydride or phthalimide in the presence of a metal salt.

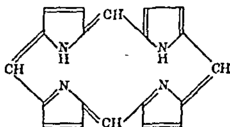
(ii) By heating *o*-cyanobenzamides or phthalonitriles with metals or metallic salts.

(iii) By heating phthalic anhydride or phthalimide with urea and a metallic salt, preferably in the presence of a catalyst such as boric acid.

Phthalocyanine, I, the parent substance of this group, may be prepared by heating phthalonitrile with a little triethanolamine. It can be seen from formula I that phthalocyanine contains four *iso*indole nuclei joined in a ring by means of nitrogen atoms. If we ignore the benzene nuclei, then we have four pyrrole nuclei linked by nitrogen atoms, a structure similar to the porphyrins,



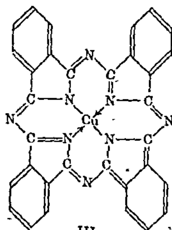
I
phthalocyanine



II
porphin.

in which the pyrrole nuclei are linked by methyne groups (II is porphin; *cf.* §2). Both types of compounds are coloured, and both contain two imino hydrogen atoms which can be replaced to form metal complexes. Because of these similarities the phthalocyanines are often known as the tetra-azaporphyrins. The first commercial

phthalocyanine pigment was **Monastral Fast Blue BS**; this is copper phthalocyanine (III).



III

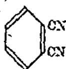
Monastral Fast Blue BS

§9. Structure of the phthalocyanines. Analysis showed that the phthalocyanines had an *empirical formula* $C_{32}H_{16}N_8M$, where M

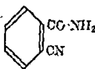
we

scopic method with naphthalene as solvent showed that the empirical formula was also the molecular formula (Linstead *et al.*, 1934). This has been confirmed by means of X-ray measurements (Robertson, Linstead *et al.*, 1935).

Linstead showed that the phthalocyanines can be obtained by reaction between a metal and phthalonitrile, I, *o*-cyanobenzamide, II, phthalamide, III, but *not* with, for example, terephthalonitrile, IV, homophthalonitrile, V, or *o*-xylylene dicyanide, VI. It is



I



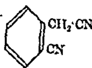
II



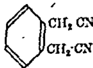
III



IV

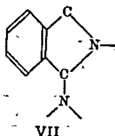


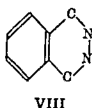
V



VI

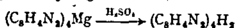
therefore reasonable to infer that in the formation of phthalocyanines, the two nitrile groups involved must be in the *ortho*-position. Thus there are probably four $C_8H_4N_2$ units, each having an *iso*indole structure, VII, or a phthalazine structure, VIII.





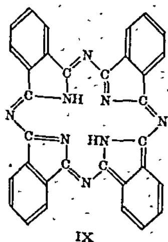
VIII was shown to be untenable since no phthalocyanine could be prepared from compounds containing this skeleton.

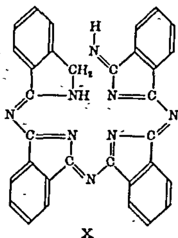
The oxidation of phthalocyanines with hot nitric acid, cold acid permanganate or ceric sulphate produces phthalimide and ammonium salts, the amount of phthalimide being that which would correspond to the presence of four *iso*indole units. The problem then is: How are these units joined together? The treatment of magnesium phthalocyanine with sulphuric acid replaces the magnesium atom by two hydrogen atoms.



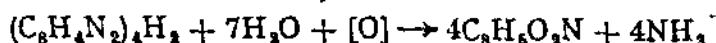
$$(C_8H_4N_2)_4Mg \xrightarrow{H_2SO_4} (C_8H_4N_2)_4H_2$$

This suggests that in metal phthalocyanines, the metal has replaced two *imino* hydrogen atoms. A reasonable structure for phthalocyanine is one in which the four *iso*indole units are joined through nitrogen atoms to form a cyclic structure (IX). On the other hand,



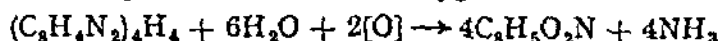


an open-chain structure could also be produced by joining four *iso*indole units through nitrogen atoms (X); in this case the molecular formula would be $(C_8H_4N_2)_4H_4$. It seems unlikely that X could be rejected on these grounds alone, since in a large molecule of this type it appears to be difficult to estimate the hydrogen with certainty (IX contains approximately 3·5 per cent. hydrogen, and X 3·9 per cent.). X, however, is unlikely, since phthalocyanine is a very stable substance; the presence of an imino group at the end of the molecule could be expected to render the compound unstable to, *e.g.*, acid reagents. Furthermore, the oxidation of phthalocyanine with ceric sulphate in dilute sulphuric acid proceeds according to the following equation (over 90 per cent. of the phthalimide has been isolated).



$$(C_8H_4N_2)_4H_2 + 7H_2O + [O] \longrightarrow 4C_8H_5O_2N + 4NH_3$$

This agrees with IX, but had the structure been X, then the molecule would have required *two* atoms of oxygen.



$$(C_8H_4N_2)_4H_4 + 6H_2O + 2[O] \longrightarrow 4C_8H_5O_2N + 4NH_3$$

Thus IX represents best the known properties of phthalocyanine. The two imino hydrogen atoms are replaceable by a bivalent metal, and the remaining two nitrogen atoms form co-ordinate links (see formula III, §8).

In metal phthalocyanines resonance is possible, and so all four nitrogen atoms linked to the metal atom would be equivalent. Phthalocyanines (with and without a central metal atom) have been examined by means of X-ray analysis (Robertson, 1936), and the results show that these compounds are large *flat* molecules with a centre of symmetry. The bond lengths of the C—N bonds indicate resonance, as do those of the benzene ring (all the lengths are equal). Robertson also showed that for nickel phthalocyanine, if the radius of the nickel atom be assumed, then the positions of the other atoms in the molecule are exactly those obtained by chemical evidence.

READING REFERENCES

Stewart and Graham, *Recent Advances in Organic Chemistry*, Longmans, Green. Vol. III (1948, 7th ed). (i) Ch. 5 Some Natural Porphyrins and Related Compounds. (ii) Ch. 6. The Azaporphyrins.

Phthalocyanines.
(i) Linstead *et al.*, *J.C.S.*, **1934**, 1016; **1936**, 1745.
(ii) Robertson, *J.C.S.*, **1935**, 615; **1936**, 1195; **1937**, 219; **1940**, 36.
(iii) Dahlen, *Ind. Eng. Chem.*, 1939, **31**, 839.

Elderfield (Ed.), *Heterocyclic Compounds*, Wiley. Vol. I (1950). *Ch. 6.* Chemistry of Pyrrole and its Derivatives

Fischer and Orth, *Die Chemie des Pyrrols*, Leipzig. Vol. II (Part I, 1937; Part II, 1940).

Gilman (Ed.), *Advanced Organic Chemistry*, Wiley (1943, 2nd ed.). (i) Ch. 16. The Chemistry of the Porphyrins. (ii) Ch. 17. Chlorophyll.

Lemberg and Legge, *Hæmatin Compounds and Bile Pigments*, Interscience Publishers (1949).

Ann. Reports (Chem. Soc.). Porphyrins. 1935, **32**, 359; 1937, **34**, 369; 1940, **37**, 318; 1950, **47**, 271.

Maitland, Biogenetic Origin of the Pyrrole Pigments, *Quart. Reviews (Chem. Soc.)*, 1950, **4**, 45.

Popják, Chemistry, Biochemistry, and Isotopic Techniques, Lectures, Monographs and Reports of the Royal Institute of Chemistry, 1955, No. 2.

Steele, Recent Progress in Determining the Chemical Structure of Chlorophyll, *Chem. Reviews*, 1937, **20**, 1.

Fischer, Chlorophyll, *Chem. Reviews*, 1937, **20**, 41.

Fischer, Chlorophyll-*a*, *J.C.S.*, 1934, 245.

Linstead, Discoveries among Conjugated Macrocyclic Compounds, *J.C.S.*, 1953, 2873.

Willstätter and Stoll, *Investigations on Chlorophyll*, Science Press (1928).

INDEX OF AUTHORS

Names associated with reactions, syntheses, etc , are not listed here , they are described in the Subject Index

K

L

M

E

F

I

N

O

U

V

